

ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ

ЗАТВЕРДЖУЮ

Декан математичного факультету ЗНУ

С.І.Гоменюк

(підпис)

(ініціали та прізвище)

« 02 » М. С. 09 2024р.

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

АНАЛІЗ ДАНИХ. ВЕЛИКІ ДАНІ

підготовки _____ магістрів _____

денної та заочної форми здобуття освіти

освітньо-професійна програма _____ Комп'ютерні науки _____

спеціальності _____ 122 Комп'ютерні науки _____

галузі знань _____ 12 Інформаційні технології _____

ВИКЛАДАЧ: Шило Галина Миколаївна, д.т.н, доцент, завідувач кафедри
комп'ютерних наук

Обговорено та ухвалено
на засіданні кафедри комп'ютерних наук

Протокол № 1 від " 29 " серпня 2024 р.
Завідувач кафедри комп'ютерних наук

Погоджено
Гарант освітньо-професійної програми



(підпис)

Г. М. Шило

(ініціали, прізвище)



(підпис)

Г. М. Шило

(ініціали, прізвище)

2024 рік

Зв'язок з викладачем: *Шило Галина Миколаївна*

E-mail: *shilo.gn@gmail.com*

СЕЗН ЗНУ повідомлення: <https://moodle.znu.edu.ua/user/profile.php?id=55852>

Телефон (кафедра): 289-12-57

Інші засоби зв'язку: Viber, Telegram

Кафедра комп'ютерних наук, ауд. №39, 1 корпус ЗНУ

1. Опис навчальної дисципліни

Дисципліна «Аналіз даних. Великі дані» надає вивчення методів сучасної обробки даних – інтелектуального аналізу даних (Data Mining, Knowledge Discovery in Data), аналітичного дослідження великих масивів інформації з метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів і різних інструментальних засобів, які використовуються Data Mining; розгляд практичних прикладів.

Метою вивчення навчальної дисципліни «Аналіз даних. Великі дані» є оволодіння знаннями, формування умінь і навичок застосування сучасних програмних засобів та методів аналізу і візуалізації даних (включно з великим) для розробки інтелектуальних систем.

У результаті вивчення навчальної дисципліни «Аналіз даних. Великі дані» студент зможе:

- створювати інфографіку та аналітичні звіти;
- обґрунтовувати вибір конкретного типу моделі та методу інтелектуального аналізу даних при вирішенні поставленої практичної задачі;
- проводити необхідну попередню обробку даних, визначати тип задачі аналізу, вирішувати її адекватно обраним методом з оптимально визначеними параметрами, оцінювати результати, робити змістовні висновки та інтерпретацію;
- використовувати сучасні програмні засоби для проектування та дослідження систем інтелектуального аналізу даних.

Компетентності, отримані під час вивчення дисципліни «Аналіз даних. Великі дані», є необхідними для вивчення дисципліни «Інтелектуальні інформаційні системи» та виконання кваліфікаційної роботи магістра.

Паспорт навчальної дисципліни

Нормативні показники	денна форма здобуття освіти	заочна форма здобуття освіти
Статус дисципліни	Обов'язкова	
Семестр	1-й	1-й
Кількість кредитів ECTS	3	
Кількість годин	90	
Лекційні заняття	10 год.	4 год.
Лабораторні заняття	20 год.	4 год.
Самостійна робота	60 год.	82 год.
Консультації	За розкладом (http://surl.li/mitciz); дистанційно: Ідентифікатор конференції Zoom: 784 7371 3154; Код доступу: 2023	
Вид підсумкового семестрового контролю:	залік	
Посилання на електронний курс у СЕЗН ЗНУ (платформа Moodle)	https://moodle.znu.edu.ua/course/view.php?id=16942	

2. Методи досягнення запланованих освітньою програмою компетентностей і результатів навчання

Компетентності/ результати навчання	Методи навчання	Форми і методи оцінювання
Компетентності		
ЗК01. Здатність до абстрактного мислення, аналізу та синтезу ЗК02. Здатність застосовувати знання у практичних ситуаціях СК4. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень СК6. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук	лекція-візуалізація, пояснення, демонстрування, дискусія, аналіз, виконання завдань лабораторних і самостійних робіт	Поточний контроль: захист лабораторних та самостійних робіт, опитування, тестування Підсумковий контроль: тестування

Компетентності/ результати навчання	Методи навчання	Форми і методи оцінювання
Результати навчання		
<p>RH1. Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань</p> <p>RH2. 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.</p> <p>RH8. Розробляти математичні моделі та методи аналізу даних (включно з великим)</p> <p>RH16. Виконувати дослідження у сфері комп'ютерних наук</p>	<p>лекція-візуалізація, пояснення, демонстрування, дискусія, аналіз, виконання завдань лабораторних і самостійних робіт</p>	<p>Поточний контроль: захист лабораторних та самостійних робіт, опитування, тестування</p> <p>Підсумковий контроль: тестування</p>

3. Зміст навчальної дисципліни

Змістовий модуль 1. Візуалізація даних

Тема 1. Основи візуалізації даних. Сучасні інструменти візуалізації даних.
 Типи даних. Мітки та візуальні канали. Вибір техніки візуалізації. Візуалізація даних у Google Data Studio. Інфодизайн: графіки та ефективність візуального кодування. Основи бізнес-аналітики. Основні поняття графічного дизайну. Основи композиції. Контраст. Візуальна вага. Рух. Ритм. Баланс. Психологічні принципи у графічному дизайні. Табличні дані і графіки. Географічні дані. Аналіз прикладів візуалізацій. Інструменти веб-аналітики. Google Analytics. Планування KPI.

Змістовий модуль 2. Класифікація

Тема 2. Основи інтелектуального аналізу даних. Задача класифікації.
 Визначення Data Mining і область застосування. Задачі, моделі та методи Data Mining. Методи, стадії, задачі Data Mining. Поняття Business Intelligence. Цикл одержання, попередньої обробки, аналізу даних, інтерпретації результатів та їхнього використання. Етапи процесу Data Mining, пов'язані з побудовою, перевіркою, оцінкою, вибором и корекцією моделей. Методи первісної обробки даних. Інструментальні засоби Data Mining. Методи дослідження структури даних: візуалізація даних.

Постановка задачі класифікації та представлення результатів. Методи побудови правил класифікації. Методи побудови дерев рішень. Методи

побудови математичних функцій. Методи опорних векторів, «найближчого сусіда», Байеса. Аналіз багатомірних угруповань. Класифікація об'єктів у випадку невідомих розподілень даних. Методи оцінювання помилок класифікації.

Змістовий модуль 3. Прогнозування

Тема 3. Прогнозування. Задача регресії. Методи аналізу часових рядів. Задача кластеризації. Сутність задачі прогнозування. Методи вирішення задачі регресії. Поняття нечітких часових рядів. Методи моделювання часових рядів. Методи аналізу та прогнозування поведінки часових рядів. Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації

Змістовий модуль 4. Пошук асоціативних правил

Тема 4. Розв'язання задачі пошуку асоціативних правил. Сховища даних та оперативний аналіз даних (OLAP). Постановка задачі пошуку асоціативних правил та представлення результатів. Секвенціальний аналіз. Різновиди задач пошуку асоціативних правил. Методи подання результатів. Алгоритми пошуку асоціативних правил. Методи пошуку асоціативних правил: метод Apriori, побудова FP-дерев пошуку шаблонів даних. Min-max асоціації у базах даних. Побудова hash-дерев.

Визначення сховища даних, порівняння з базами даних, використання. Архітектура сховища даних. ETL-процеси (добування, перетворення й завантаження даних). Вітрини даних, куби даних, багатомірна модель даних. Розгортання OLAP-кубів. Операції над OLAP-кубами (зріз, обертання, консолідація, деталізація). Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.

Змістовий модуль 5. Великі дані

Тема 5 Аналіз та обробка великих даних. Введення до Apache Hadoop, Hadoop Distributed File System та MapReduce. Введення до Apache Spark. Поняття Resilient Distributed Dataset

4. Структура навчальної дисципліни

Вид заняття / роботи	Назва теми	Кількість годин		Згідно з розкладом
		о/д. ф.	з.ф.	
1	2	3	4	5
Лекція 1	Тема. Основи візуалізації даних. Сучасні інструменти візуалізації даних. Інфодизайн: графіки та ефективність візуального кодування. Основи бізнес-аналітики	2	0,5	тиждень 1
Лабораторна робота 1	Візуалізація даних у Google Data Studio	2	-	тиждень 1
Лабораторна робота 2	Створення аналітичного звіту. Веб-аналітика	2	0,5	тиждень 2

Вид заняття /роботи	Назва теми	Кількість годин		Згідно з розкладом
		о/д. ф.	з.ф.	
Самостійна робота	Індивідуальне завдання	15	20	тиждень 1,2
Лекція 2	Основи інтелектуального аналізу даних. Задача класифікації	2	0,5	тиждень 3
Лабораторна робота 3	Підготовка даних для інтелектуального аналізу даних з використанням Weka	2	0,5	тиждень 3
Лабораторна робота 4	Задача класифікації	2	0,5	тиждень 4
Самостійна робота	Індивідуальне завдання	15	20	тижні 3,4
Лекція 3	Прогнозування. Задача регресії. Методи аналізу часових рядів. Задача кластеризації	2	1	тиждень 5
Лабораторна робота 5	Прогнозування. Задача регресії	2	0,5	тиждень 5
Лабораторна робота 6	Задача кластеризації	2	0,5	тиждень 6
Самостійна робота	Індивідуальне завдання	15	22	тиждень 5,6
Лекція 4	Розв'язання задачі пошуку асоціативних правил. Сховища даних та оперативний аналіз даних (OLAP)	2	1	тиждень 7
Лабораторна робота 7	Пошук асоціативних правил	4	0,5	тиждень 7,8
Самостійна робота	Індивідуальне завдання	15	20	тиждень 7,8
Лекція 5	Аналіз та обробка великих даних	2	1	тиждень 9
Лабораторна робота 8	Програмне забезпечення обробки великих даних Apache Hadoop, MapReduce	2	0,5	тиждень 9
Лабораторна робота 9	Програмне забезпечення обробки великих даних Apache Spark	2	0,5	тиждень 10

Методичні рекомендації до лабораторних та самостійних занять розміщено СЕЗН ЗНУ Moodle на сторінці дисципліни.

5. Види і зміст контрольних заходів

Вид заняття/ роботи	Вид контрольного заходу	Зміст контрольного заходу	Критерії оцінювання та термін виконання	Усього балів
Поточний контроль				
Лабораторна робота №1	Захист лабораторної роботи 1	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування, тиждень 1	5

Запорізький національний університет
Силабус навчальної дисципліни
Аналіз даних. Великі дані



Вид заняття/ роботи	Вид контрольного заходу	Зміст контрольного заходу	Критерії оцінювання та термін виконання	Усього балів
Лабораторна робота №2	Захист лабораторної роботи 2	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування, тиждень 2	5
Лабораторна робота №3	Захист лабораторної роботи 3	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування, тиждень 3	5
Лабораторна робота №4	Захист лабораторної роботи 4	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування, тиждень 4	5
Лабораторна робота №5	Захист лабораторної роботи 5	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування, тиждень 5	5
Лабораторна робота №6	Захист лабораторної роботи 6	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування, тиждень 6	5
Лабораторна робота №7	Захист лабораторної роботи 7	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування, тиждень 8	5
Лабораторна робота №8	Захист лабораторної роботи 8	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування, тиждень 9	5
Лабораторна робота №9	Захист лабораторної роботи 9	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування, тиждень 10	5
Самостійні роботи	Захист індивідуального завдання	Розміщено в СЕЗН ЗНУ	Виконання та завантаження пояснювальної записки у СЕЗН ЗНУ, усне опитування, тиждень 10	15
Усього за поточний контроль	10			60
Підсумковий контроль				
Залік	Тестування	Розміщено в СЕЗН ЗНУ	20 тестових завдань правильна відповідь на кожне оцінюється в 2 бали	40
Усього за підсумковий контроль				40

Шкала оцінювання ЗНУ: національна та ECTS

За шкалою ECTS	За шкалою університету	За національною шкалою	
		Екзамен	Залік
A	90 – 100 (відмінно)	5 (відмінно)	Зараховано
B	85 – 89 (дуже добре)	4 (добре)	
C	75 – 84 (добре)		
D	70 – 74 (задовільно)		
E	60 – 69 (достатньо)	3 (задовільно)	
FX	35 – 59 (незадовільно – з можливістю повторного складання)	2 (незадовільно)	Не зараховано
F	1 – 34 (незадовільно – з обов’язковим повторним курсом)		

6. Основні навчальні ресурси

Рекомендована література

- Муляр В. П. Візуалізація даних та інфографіка. Харків : ФОП Панов А. М., 2020. 200 с.
- Гороховатський В. О., Творошенко І. С. Методи інтелектуального аналізу та оброблення даних : навч. посібник. Харків : ХНУРЕ, 2021. 92 с.
- Лупан І. В. Інтелектуальний аналіз даних Data Mining : навчально-методичний посібник. Кропивницький : ФОП Піскова М. А., 2022. 112 с.
- Сергеев-Горчинський О. О., Іщенко Г. В. Інтелектуальний аналіз даних : комп'ютерний практикум. Електронні текстові данні (1файл: 1,72 Мбайт). Київ : КПІ ім. Ігоря Сікорського, 2018. 73 с.:
- Балабанов О. С. Аналітика великих даних: принципи, напрямки і задачі (огляд). *Проблеми програмування*. 2019. № 2. С. 47–68.
- Балабанов О. С. Задачі та методи аналізу великих даних : (огляд). *Проблеми програмування*. 2019. № 3. С. 58–85.
- Ланде Д. В., Субач І. Ю., Гладун А. Я. Оброблення надвеликих масивів даних (Big Data) : навч. посіб. для використання у навч. процесі з підготовки фахівців другого (магістерського) рівня вищої освіти зі спец. 122 «Комп'ютерні науки». Київ : ІСЗІ КПІ ім. Ігоря Сікорського, 2021. 168 с.
- Олещенко Л. М. Технології оброблення великих даних : конспект лекцій : навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2021. 227 с.
- Олещенко Л. М. Технології оброблення великих даних: комп'ютерний практикум : навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2021. 85 с.
URL: <http://files.znu.edu.ua/files/Bibliobooks/Inshi72/0053478.pdf>.

10. Таран В. І., Гордієнко Ю. Г., Стіренко С. Г. Технології BigData : навч. посіб. для здобувачів ступеня магістра. Київ : КПІ ім. Ігоря Сікорського, 2022. 56 с.
11. Berman J. J. Principles and Practice of Big Data : Preparing, Sharing, and Analyzing Complex Information. 2nd ed. Cambridge : Elsevier, 2018. 480 p. URL: <http://ebooks.znu.edu.ua/files/Bibliobooks/ScienceDirect/0046085.zip>
12. Intelligent Computing for Big Data / W. Wang, Ka Lok Man (eds.). Basel : MDPI, 2022. 100 p.
13. Kalita J. K., Bhattacharyya D. K., Roy S. Fundamentals of Data Science: Theory and Practice. London : Academic Press, 2024. 307 p.

Інформаційні ресурси

1. Weka 3: Machine Learning Software in Java. URL: <http://www.cs.waikato.ac.nz/ml/weka/>

7. Регуляції і політики курсу

Відвідування занять. Регуляція пропусків.

Відвідування усіх занять є обов'язковим. Студенти зобов'язані дотримуватися усіх строків, визначених для виконання усіх видів робіт, передбачених даною дисципліною. Пропуски та запізнення на заняття є недопустимими.

Політика академічної доброчесності

Кожний студент зобов'язаний дотримуватися принципів академічної доброчесності. Письмові завдання з використанням часткових або повнотекстових запозичень з інших робіт без зазначення авторства – це *плагіат*. Використання будь-якої інформації (текст, фото, ілюстрації тощо) мають бути правильно процитовані з посиланням на автора! Якщо ви не впевнені, що таке плагіат, фабрикація, фальсифікація, порадьтеся з викладачем. До студентів, у роботах яких буде виявлено списування, плагіат чи інші прояви недоброчесної поведінки можуть бути застосовані різні дисциплінарні заходи (див. посилання на Кодекс академічної доброчесності ЗНУ в додатку до силабусу). Неприпустиме складання роботи, виконаної іншою особою.

Використання комп'ютерів/телефонів на занятті

Використання мобільних телефонів, ноутбуків та інших гаджетів під час лекційних та лабораторних занять дозволяється виключно у навчальних цілях (з активованим режимом «без звуку»).

Комунікація

Комунікація викладача зі студентами здійснюється безпосередньо на заняттях та додатково за допомогою месенджерів (наприклад, Telegram), електронної пошти і в СЕЗН Moodle (форум курсу, приватні повідомлення)

ДОДАТКОВА ІНФОРМАЦІЯ

ГРАФІК ОСВІТНЬОГО ПРОЦЕСУ 2024-2025 н. р. доступний за адресою:
<https://tinyurl.com/yckze4jd>.

НАВЧАЛЬНИЙ ПРОЦЕС ТА ЗАБЕЗПЕЧЕННЯ ЯКОСТІ ОСВІТИ. Перевірка набутих студентами знань, навичок та вмінь (атестації, заліки, іспити та інші форми контролю) є невід'ємною складовою системи забезпечення якості освіти і проводиться відповідно до Положення про організацію та методику проведення поточного та підсумкового семестрового контролю навчання студентів ЗНУ: <https://tinyurl.com/y9tve4lk>.

ПОВТОРНЕ ВИВЧЕННЯ ДИСЦИПЛІН, ВІДРАХУВАННЯ. Наявність академічної заборгованості до 6 навчальних дисциплін (в тому числі проходження практики чи виконання курсової роботи) за результатами однієї екзаменаційної сесії є підставою для надання студенту права на повторне вивчення зазначених навчальних дисциплін. Порядок повторного вивчення визначається Положенням про порядок повторного вивчення навчальних дисциплін та повторного навчання у ЗНУ: <https://tinyurl.com/y9pkmp5>. Підстави та процедури відрахування студентів, у тому числі за невиконання навчального плану, регламентуються Положенням про порядок переведення, відрахування та поновлення студентів у ЗНУ: <https://tinyurl.com/ycds57la>.

ВИРІШЕННЯ КОНФЛІКТІВ. Порядок і процедури врегулювання конфліктів, пов'язаних із корупційними діями, зіткненням інтересів, різними формами дискримінації, сексуальними домаганнями, міжособистісними стосунками та іншими ситуаціями, що можуть виникнути під час навчання, регламентуються Положенням про порядок і процедури вирішення конфліктних ситуацій у ЗНУ: <https://tinyurl.com/57wha734>. Конфліктні ситуації, що виникають у сфері стипендіального забезпечення здобувачів вищої освіти, вирішуються стипендіальними комісіями факультетів, коледжів та університету в межах їх повноважень, відповідно до: Положення про порядок призначення і виплати академічних стипендій у ЗНУ: <https://tinyurl.com/yd6bq6p9>; Положення про призначення та виплату соціальних стипендій у ЗНУ: <https://tinyurl.com/y9r5dpwh>.

ПСИХОЛОГІЧНА ДОПОМОГА. Телефон довіри практичного психолога **Марти Ірини Вадимівни** (061) 228-15-84, (099) 253-78-73 (щоденно з 9 до 21).

УПОВНОВАЖЕНА ОСОБА З ПИТАНЬ ЗАПОБІГАННЯ ТА ВИЯВЛЕННЯ КОРУПЦІЇ Запорізького національного університету: **Банах Віктор Аркадійович**
Електронна адреса: v_banakh@znu.edu.ua
Гаряча лінія: тел. (061) 227-12-76, факс 227-12-88

РІВНІ МОЖЛИВОСТІ ТА ІНКЛЮЗИВНЕ ОСВІТНЄ СЕРЕДОВИЩЕ. Центральні входи усіх навчальних корпусів ЗНУ обладнані пандусами для забезпечення доступу осіб з інвалідністю та інших маломобільних груп населення. Допомога для здійснення входу у разі потреби надається черговими охоронцями навчальних корпусів. Якщо вам потрібна спеціалізована допомога, будь ласка,

зателефонуйте (061) 228-75-11 (начальник охорони). Порядок супроводу (надання допомоги) осіб з інвалідністю та інших маломобільних груп населення у ЗНУ: <https://tinyurl.com/ydhcsagx>.

РЕСУРСИ ДЛЯ НАВЧАННЯ

НАУКОВА БІБЛІОТЕКА: <http://library.znu.edu.ua>. Графік роботи абонементів: понеділок-п'ятниця з 08.00 до 16.00; вихідні дні: субота і неділя.

СИСТЕМА ЕЛЕКТРОННОГО ЗАБЕЗПЕЧЕННЯ НАВЧАННЯ (MOODLE):
<https://moodle.znu.edu.ua>

Якщо забули пароль/логін, направте листа з темою «Забув пароль/логін» за адресою: moodle.znu@znu.edu.ua.

У листі вкажіть: прізвище, ім'я, по-батькові українською мовою; шифр групи; електронну адресу.

Якщо ви вказували електронну адресу в профілі системи Moodle ЗНУ, то використовуйте посилання для відновлення паролю
<https://moodle.znu.edu.ua/mod/page/view.php?id=133015>.

ЦЕНТР ІНТЕНСИВНОГО ВИВЧЕННЯ ІНОЗЕМНИХ МОВ:
<http://sites.znu.edu.ua/child-advance/>

ЦЕНТР НІМЕЦЬКОЇ МОВИ, ПАРТНЕР ГЕТЕ-ІНСТИТУТУ:
<https://www.znu.edu.ua/ukr/edu/ocznu/nim>

ШКОЛА КОНФУЦІЯ (ВИВЧЕННЯ КИТАЙСЬКОЇ МОВИ):
<http://sites.znu.edu.ua/confucius>