

ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ

ЗАТВЕРДЖУЮ

Декан математичного факультету

С.І. Гоменюк

«02» вересня 2024 р.

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

АНАЛІЗ ДАНИХ. ВЕЛИКІ ДАНІ

підготовки магістрів

денної форми здобуття освіти

освітньо-наукова програма «Комп'ютерні науки»

спеціальності 122 Комп'ютерні науки

галузі знань 12 Інформаційні технології

ВИКЛАДАЧ: Шило Г. М., д.т.н., професор, 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



Шило Г. М.

Погоджено

Гарант освітньо-наукової програми



С. І. Гоменюк

Зв'язок з викладачем: Шило Галина Миколаївна

E-mail: shilo.gn@gmail.com

СЕЗН ЗНУ повідомлення: <https://moodle.znu.edu.ua/course/view.php?id=16942>

Телефон (кафедра): 289-12-57

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

1. Опис навчальної дисципліни

Дисципліна «Аналіз даних. Великі дані» надає можливість ознайомитися з сучасними концепціями машинного навчання, основами алгоритмів, що використовуються для аналізу даних, а також практичними методами їх застосування в різних сферах комп'ютерних наук.

Метою вивчення навчальної дисципліни «Аналіз даних. Великі дані» є оволодіння основними техніками та підходами в машинному навчанні, а також розвиток навичок самостійного вирішення практичних завдань за допомогою машинного навчання.

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

- проектувати та реалізовувати моделі машинного навчання для розв'язання реальних задач;
- працювати з великими обсягами даних, включаючи їх підготовку та попередню обробку;
- обирати та обґрунтовувати методи машинного навчання в залежності від специфіки задачі;
- оцінювати ефективність моделей і здійснювати їх вдосконалення.

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

Паспорт навчальної дисципліни

Нормативні показники	денна форма здобуття освіти
Статус дисципліни	Обов'язкова
Семестр	1-й
Кількість кредитів ECTS	3
Кількість годин	90
Лекційні заняття	10 год.
Лабораторні заняття	20 год.
Самостійна робота	60 год.
Консультації	За розкладом; дистанційно
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 практичних робіт	Поточний контроль: захист лабораторних та самостійних робіт, опитування, тестування Підсумковий контроль: тестування

Компетентності/ результати навчання	Методи навчання	Форми і методи оцінювання
Результати навчання		
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 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. RH8. Розробляти математичні моделі та методи аналізу даних (включно з великим) RH9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими) RH12. Проектувати та супроводжувати бази даних та знань RH16. Виконувати дослідження у сфері комп'ютерних наук RH20. Створювати та досліджувати інформаційні та математичні моделі систем і процесів, що досліджуються, зокрема об'єктів автоматизації	лекція-візуалізація, пояснення, дискусія, аналіз, виконання завдань практичних робіт	Поточний контроль: захист лабораторних та самостійних робіт, опитування, тестування Підсумковий контроль: тестування

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 о/д. ф.	Згідно з розкладом
Лекція 1	Тема. Основи візуалізації даних. Сучасні інструменти візуалізації даних. Інфодизайн: графіки та ефективність візуального кодування. Основи бізнес-аналітики	2	тиждень 1
Лабораторна робота 1	Візуалізація даних у Google Data Studio	2	тиждень 1
Лабораторна робота 2	Створення аналітичного звіту. Веб-аналітика	2	тиждень 2
Самостійна робота	Індивідуальне завдання	15	тиждень 1,2
Лекція 2	Основи інтелектуального аналізу даних. Задача класифікації	2	тиждень 3
Лабораторна робота 3	Підготовка даних для інтелектуального аналізу даних з використанням Weka	2	тиждень 3
Лабораторна робота 4	Задача класифікації	2	тиждень 4
Самостійна робота	Індивідуальне завдання	15	тижні 3,4
Лекція 3	Прогнозування. Задача регресії. Методи аналізу часових рядів. Задача кластеризації	2	тиждень 5
Лабораторна робота 5	Прогнозування. Задача регресії	2	тиждень 5
Лабораторна робота 6	Задача кластеризації	2	тиждень 6
Самостійна робота	Індивідуальне завдання	15	тиждень 5,6
Лекція 4	Розв'язання задачі пошуку асоціативних правил. Сховища даних та оперативний аналіз даних (OLAP)	2	тиждень 7
Лабораторна робота 7	Пошук асоціативних правил	4	тиждень 7,8
Самостійна робота	Індивідуальне завдання	15	тиждень 7,8
Лекція 5	Аналіз та обробка великих даних	2	тиждень 9
Лабораторна робота 8	Програмне забезпечення обробки великих даних Apache Hadoop, MapReduce	2	тиждень 9
Лабораторна робота 9	Програмне забезпечення обробки великих даних Apache Spark	2	тиждень 10

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

5. Види і зміст контрольних заходів

Вид заняття/ роботи	Вид контрольного заходу	Зміст контрольно го 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	5
Лабораторна робота №2	Захист лабораторної роботи 2	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування	5
Лабораторна робота №3	Захист лабораторної роботи 3	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування	5
Лабораторна робота №4	Захист лабораторної роботи 4	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування	5
Лабораторна робота №5	Захист лабораторної роботи 5	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування	5
Лабораторна робота №6	Захист лабораторної роботи 6	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування	5
Лабораторна робота №7	Захист лабораторної роботи 7	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування	5
Лабораторна робота №8	Захист лабораторної роботи 8	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування	5
Лабораторна робота №9	Захист лабораторної роботи 9	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування	5
Самостійні роботи	Захист індивідуального завдання	Розміщено в СЕЗН ЗНУ	Виконання та завантаження звіту у СЕЗН ЗНУ, усне опитування	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)	3 (задовільно)	
E	60 – 69 (достатньо)		
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

1. Муляр В. П. Візуалізація даних та інфографіка. Харків : ФОП Панов А. М., 2020. 200 с.
2. Гороховатський В. О., Творошенко І. С. Методи інтелектуального аналізу та оброблення даних : навч. посібник. Харків : ХНУРЕ, 2021. 92 с.
3. Лупан І. В. Інтелектуальний аналіз даних Data Mining : навчально-методичний посібник. Кропивницький : ФОП Піскова М. А., 2022. 112 с.
4. 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.:
5. Балабанов О. С. Аналітика великих даних: принципи, напрямки і задачі (огляд). Проблеми програмування. 2019. № 2. С. 47–68.
6. Балабанов О. С. Задачі та методи аналізу великих даних : (огляд). Проблеми програмування. 2019. № 3. С. 58–85.
7. 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.
8. Олещенко Л. М. Технології оброблення великих даних : конспект лекцій : навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2021. 227 с.
9. 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

14. 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, 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/y9pkmm5>. 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>