

Зв'язок з викладачем: Добровольський Геннадій Анатолійович

E-mail: gen.dobr@gmail.com

Сезн ЗНУ повідомлення: <https://moodle.znu.edu.ua/enrol/index.php?id=16922>

Телефон (кафедра): +38-061-289-12-57

Інші засоби зв'язку: https://t.me/gen_dobr

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

1. Опис навчальної дисципліни

Метою викладання навчальної дисципліни «Аналіз даних. Великі дані» є знайомство студентів з основними концепціями, технологіями та методами обробки великих даних (Big Data), їхньою архітектурою, аналітикою та застосуванням у реальних сценаріях. Компетентності, отримані під час вивчення дисципліни «Аналіз даних. Великі дані», необхідні для виконання завдань виробничої практики та кваліфікаційної роботи магістра.

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

- створювати конвеєри обробки та аналізу даних, використовуючи Apache Spark;
- створювати системи зберігання великих даних;
- створювати різноманітні візуальні представлення даних за допомогою платформи Tableau.

Паспорт навчальної дисципліни

Нормативні показники	денна форма здобуття освіти	заочна форма здобуття освіти
Статус дисципліни	Вибіркова	
Семестр	3-й	3-й
Кількість кредитів ECTS	4	4
Кількість годин	120	120
Лекційні заняття	12 год.	4 год.
Практичні заняття	22 год.	4 год.
Самостійна робота	86 год.	112 год.
Консультації	Згідно розкладу: дистанційно: Zoom: Ідентифікатор: 771 5037 0584 Пароль: 9K8yqV	
Вид підсумкового семестрового контролю:	залік	
Посилання на електронний курс у СЕЗН ЗНУ (платформа Moodle)	https://moodle.znu.edu.ua/enrol/index.php?id=16922	

2. Методи досягнення запланованих освітньою програмою компетентностей і результатів навчання

Компетентності/ результати навчання	Методи навчання	Форми і методи оцінювання
Компетентності		
<ul style="list-style-type: none"> - Здатність до абстрактного мислення, аналізу та синтезу - Здатність застосовувати знання у практичних ситуаціях - Здатність спілкуватися державною мовою як усно, так і письмово - Здатність вчитися й оволодівати сучасними знаннями - Здатність бути критичним і самокритичним - Здатність генерувати нові ідеї (креативність) - Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі 	лекція-візуалізація, пояснення, демонстрування, дискусія, аналіз, виконання завдань практичних робіт	Поточний контроль: захист практичних та самостійних робіт, опитування, тестування Підсумковий контроль: тестування

<ul style="list-style-type: none"> - Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проектних рішень - Здатність розробляти і реалізовувати проекти зі створення програмного забезпечення, у тому числі в непередбачуваних умовах, за нечітких вимог та необхідності застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проектом 		
Результати навчання		
<ul style="list-style-type: none"> - Зрозуміло і недвозначно доносити власні знання, висновки та аргументацію у сфері комп'ютерних наук до фахівців і нефахівців, зокрема до осіб, які навчаються - Розробляти та застосовувати математичні методи для аналізу інформаційних моделей - Розробляти математичні моделі та методи аналізу даних (включно з великим) 	<p>лекція-візуалізація, пояснення, дискусія, аналіз, виконання завдань практичних робіт</p>	<p>Поточний контроль: захист практичних та самостійних робіт, опитування, тестування</p> <p>Підсумковий контроль: тестування</p>

3. Зміст навчальної дисципліни

Змістовий модуль 1. Вступ до великих даних

Тема 1. Вступ до великих даних.

Що таке великі дані. Огляд 5 V: Volume, Variety, Velocity, Veracity, Value. Сучасні проблеми й виклики обробки даних у великих масштабах. Огляд застосувань великих даних у бізнесі, науці та техніці.

Змістовий модуль 2. Зберігання великих даних.

Тема 2. Зберігання великих даних

Традиційні бази даних vs великі дані. Хмарні платформи для обробки великих даних. Архітектура обробки даних: ETL-процеси, Data Lake, Data

Warehouse. Моделі зберігання та розподілені файлові системи (Hadoop HDFS). Порівняння реляційних і NoSQL баз даних. Типи NoSQL баз даних: Key-Value, Column-Oriented, Document-Oriented, Graph Databases. Огляд популярних NoSQL систем: MongoDB, Cassandra, Hbase.

Змістовий модуль 3. Hadoop, Apache Spark

Тема 3. Огляд систем Hadoop та Apache Spark

Компоненти Hadoop: HDFS, YARN, MapReduce. Вступ до MapReduce: основи програмування й оптимізація. Налаштування кластера Hadoop і написання простих MapReduce програм. Apache Spark: основи та порівняння з MapReduce. RDD, DataFrames, SQL.

Тема 4. Інструменти обробки великих даних

Інструменти обробки та аналітики великих даних. Огляд Apache Hive та Pig. Написання запитів на Apache Hive для обробки великих обсягів даних. Хмарні інструменти обробки та аналітики великих даних.

Змістовий модуль 4. Потокова обробка даних

Тема 5. Потокова обробка даних в Apache Spark

Потоки даних. Джерела потоків. Мікропакети. Structured Streaming. Трансформації потоку. Агрегації та вікна. Обробка із збереженням стану.

Тема 6. Потокова обробка даних в Apache Kafka.

Apache Kafka: архітектура та принципи роботи. Налаштування та використання Apache Kafka для потокової передачі даних.

Змістовий модуль 5. Машинне навчання на великих даних

Тема 7. Машинне навчання на великих даних в Apache Spark.

Використання машинного навчання для великих даних. Огляд бібліотек для машинного навчання в екосистемах Hadoop і Spark (Mllib). Обробка великих наборів даних для навчання моделей. Побудова моделей машинного навчання з Apache Spark MLlib.

Змістовий модуль 6. Застосування засобів візуальної аналітики даних

Тема 8. Застосування засобів візуальної аналітики даних

Роль візуальної аналітики у вирішенні складних проблем та обробці великих обсягів інформації. Основні завдання візуальної аналітики. Основні принципи візуальної аналітики: простота, чіткість, адекватність даним. Приклади застосування візуальної аналітики.

Тема 9. Платформа Tableau.

Огляд платформи Tableau. Основні можливості: швидкий аналіз великих наборів даних, інтерактивні дашборди, гнучкість у побудові візуалізацій. Приклади застосування. Версії Tableau: Tableau Desktop, Tableau Public, Tableau Server, Tableau Online. Основні елементи інтерфейсу Tableau. Виміри (Dimensions) та показники (Measures). Категоріальні та кількісні дані. Підключення до даних та функції роботи з даними. Типи візуалізацій у Tableau: лінійні графіки, гістограми, стовпчикові діаграми, діаграми розсіювання, теплові карти, кругові діаграми. Створення інтерактивного дашборду.

4. Структура навчальної дисципліни

Вид заняття /роботи	Назва теми	Кількість годин о/д. ф.	Згідно з розкладом
1	2	3	4
Лекція 1	Вступ до великих даних.	2	тиждень 1
Практичне заняття 1	Огляд застосувань великих даних у бізнесі, науці та техніці.	2	тиждень 1
Самостійна робота	Критерії якості даних та їх оцінка	14	тиждень 1
Лекція 2	Зберігання великих даних	2	тиждень 2
Практичне заняття 2	Робота з MongoDB (створення, читання, оновлення, видалення даних).	4	тиждень 2 тиждень 3
Самостійна робота	ETL-процеси, Data Lake, Data Warehouse	15	тиждень 2 тиждень 3
Лекція 3	Огляд систем Hadoop та Apache Spark	2	тиждень 4
Практичне заняття 3	Написання програм на Apache Spark (RDD, DataFrames, SQL).	4	тиждень 4 тиждень 5
Самостійна робота	Хмарні інструменти обробки та аналітики великих даних.	15	тиждень 4 тиждень 5
Лекція 4	Потокова обробка даних	2	тиждень 6
Практичне заняття 4	Потокова обробка даних в Apache Spark	4	тиждень 6 тиждень 7
Самостійна робота	Використання Apache Kafka для потокової передачі даних.	14	тиждень 6 тиждень 7
Лекція 5	Машинне навчання на великих даних в Apache Spark	2	тиждень 8
Практичне заняття 5	Побудова моделей машинного навчання з Apache Spark MLlib	4	тиждень 8 тиждень 9
Самостійна робота	Хмарні інструменти машинного навчання.	14	тиждень 8 тиждень 9
Лекція 6	Візуалізація великих даних	2	тиждень 10
Практичне заняття 6	Створення процесу аналітики даних у Tableau.	4	тиждень 10 тиждень 11
Самостійна робота	Уроки з документації Tableau. Принципи ефективної інфографіки, оцінка ефективності	14	тиждень 10 тиждень 11

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 контрольного заходу*	Критерії оцінювання та термін виконання*	Усього балів
Поточний контроль				
Практичне заняття №1	Практична робота 1	Розміщено в СЕЗН ЗНУ	Повний захист виконаного завдання оцінюється в 8 балів, тільки зданий звіт - 6 балів	8
Практичне заняття №2	Практична робота 2	Розміщено в СЕЗН ЗНУ	Повний захист виконаного завдання оцінюється в 8 балів, тільки зданий звіт - 6 балів	8
Практичне заняття №3	Практична робота 3 Самостійна робота	Розміщено в СЕЗН ЗНУ	Повний захист виконаного завдання оцінюється в 8 балів, тільки зданий звіт - 6 балів	8
	Тест 1	Розміщено в СЕЗН ЗНУ	12 питань } по 0,5 балів	6
Практичне заняття №4	Практична робота 4	Розміщено в СЕЗН ЗНУ	Повний захист виконаного завдання оцінюється в 8 балів, тільки зданий звіт - 6 балів	8
Практичне заняття №5	Практична робота 5	Розміщено в СЕЗН ЗНУ	Повний захист виконаного завдання оцінюється в 8 балів, тільки зданий звіт - 6 балів	8
Практичне заняття №6	Практична робота 6	Розміщено в СЕЗН ЗНУ	Повний захист виконаного завдання оцінюється в 8 балів, тільки зданий звіт - 6 балів	8
	Тест 2	Розміщено в СЕЗН ЗНУ	12 питань } по 0,5 балів	6
Усього за поточний контроль				60
Підсумковий контроль				
Залік	Теоретичне завдання	Розміщено в СЕЗН ЗНУ	40 тестових питань (вибір правильної відповіді з декількох можливих) } по 1 балу	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. 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. URL: <http://files.znu.edu.ua/files/Bibliobooks/Inshi80/0059774.pdf>.
2. Олещенко Л. М. Технології оброблення великих даних : конспект лекцій : навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2021. 227 с. URL: <http://files.znu.edu.ua/files/Bibliobooks/Inshi72/0053281.pdf>.
3. 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. URL: <http://files.znu.edu.ua/files/Bibliobooks/Inshi78/0058664.pdf>.
4. Algorithms for Big Data : DFG Priority Program 1736 / Н. Bast [et al.] (eds.). Cham : Springer, 2022. 285 p. URL: <http://files.znu.edu.ua/files/Bibliobooks/Inshi70/0051051.pdf>.
5. 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>.
6. Big Data Security / S. Gupta, I. Banerjee, S. Bhattacharyya (eds.). Berlin : De Gruyter, 2019. 144 p. URL: <http://ebooks.znu.edu.ua/files/Bibliobooks/Inshi61/0045537.pdf>.
7. Cognitive Big Data Intelligence with a Metaheuristic Approach / edited by S. Mishra [et al.]. London : Academic Press, 2022. 356 p. URL: <http://files.znu.edu.ua/files/Bibliobooks/Inshi70/0051084/>.

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, порадьтесь 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/y9pkmmp5>. 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>