

**ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ»
ФАКУЛЬТЕТ МАТЕМАТИКИ ТА ЦИФРОВИХ ТЕХНОЛОГІЙ
Кафедра кібернетики і прикладної математики**



«ЗАТВЕРДЖУЮ»

Декан факультету математики

та цифрових технологій

Микола МАЛЯР/

2023 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

АНАЛІТИКА ВЕЛИКИХ ДАНИХ

Рівень вищої освіти	другий (магістерський)
Галузь знань	11 Математика та статистика
Спеціальність	113 Прикладна математика
Освітня програма	Науки про дані та інтелектуальні рішення
Статус дисципліни	обов'язкова
Мова навчання	українська

Робоча програма навчальної дисципліни «Аналітика великих даних» для здобувачів вищої освіти галузі знань **11 Математика та статистика** спеціальності **113 Прикладна математика** освітньої програми **Науки про дані та інтелектуальні рішення**.

Розробник: Млавець Ю.Ю., доцент, к.ф.-м.н., доцент кафедри кібернетики і прикладної математики.

Робочу програму розглянуто та затверджено на засіданні кафедри **кібернетики і прикладної математики**.

Протокол № 12 від «05» червня 2023 року.

Завідувач кафедри  Павло МУЛЕСА

Схвалено науково-методичною комісією **факультету математики та цифрових технологій**.

Протокол № 10 від «20» червня 2023 року.

Голова науково-методичної комісії  Наталія ЮРЧЕНКО

1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Розподіл годин за навчальним планом
	Денна форма навчання
Кількість кредитів ЄКТС – 4	Рік підготовки:
Загальна кількість годин – 120	1-й
Кількість модулів – 2	Семестр:
Тижневих годин для денної форми навчання: аудиторних – 3 самостійної роботи студента – 4	1-й
	Лекції:
	24
	Практичні (семінарські):
	-
Вид підсумкового контролю: екзамен.	Лабораторні:
	24
Форма підсумкового контролю: письмовий.	Самостійна робота:
	72

2. МЕТА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Курс «Аналітика великих даних» сьогодні є однією з найбільш затребуваних у сучасному бізнесі. Знання нових технологій програмування та уміння розробляти програмне забезпечення для управління та аналізу великих об'ємів інформації використовуються для забезпечення цифровізації усіх сфер життя.

Володіння сучасними інструментальними засобами дає змогу зосередитися на алгоритмі розв'язання практичної задачі та використовувати існуючі пакети алгоритмів для швидкого та ефективного розв'язання поставленої задачі.

Предметом дисципліни є теоретичні та практичні основи оброблення великих даних.

Мета дисципліни: сформувати теоретичні знання та практичні навички в області обробки великих даних у студентів, які можуть бути використані при подальшому навчанні, професійній, виробничій та науковій діяльності.

Студенти повинні опанувати принципи роботи з мовою та середовищем програмування R та набути практичного досвіду розв'язання практичних задач.

Вивчення дисципліни сприяє формуванню у здобувачів вищої освіти таких компетентностей:

загальні компетентності:

- здатність до абстрактного мислення, аналізу і синтезу (ЗК01);
- здатність до самонавчання, пошуку, оброблення та аналізу інформації з різних джерел (ЗК02);
- здатність генерувати нові ідеї (креативність) (ЗК04);
- здатність прийняття обґрунтованих рішень, представлення та донесення знань та ідей до широкого загалу (ЗК06).

фахові компетентності:

- здатність використовувати математичний апарат, розробляти моделі для розв'язання задач широкого спектру (ФК01);
- здатність обирати та застосовувати математичні методи для розв'язання прикладних задач, моделювання, прогнозування, прийняття рішень, аналізу даних (ФК02);
- здатність до проведення математичного і комп'ютерного моделювання, аналізу та обробки даних, обчислювального експерименту, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів (ФК03);
- здатність розробляти нові та адаптовувати вже існуючі методи та алгоритми розв'язання прикладних задач моделювання та аналізу даних, проводити відповідні експерименти з аналізом одержаних результатів (ФК04);
- здатність досліджувати наукові проблеми за фахом (ФК06);
- здатність пропонувати практичні рішення за фахом з урахуванням сучасних досягнень науки (ФК07).

3. ПЕРЕДУМОВИ ДЛЯ ВИВЧЕННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Згідно з освітньою програмою «Науки про дані та інтелектуальні рішення» передумов для вивчення навчальної дисципліни «Аналітика великих даних» не має.

4. ОЧІКУВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ

Відповідно до освітньої програми «Науки про дані та інтелектуальні рішення», вивчення навчальної дисципліни повинно забезпечити досягнення здобувачами вищої освіти таких програмних результатів навчання:

Програмні результати навчання	Шифр ПРН
Використовувати й адаптувати математичні теорії та моделі для забезпечення теоретичного підґрунтя розв'язання наукових та практичних задач.	ПРН 01

Обирати, застосовувати та розробляти нові методи й алгоритми аналізу даних для розв'язання наукових та прикладних задач.	ПРН 02
Проводити математичне і комп'ютерне моделювання, аналіз та обробку даних, обчислювальний експеримент, розв'язання формалізованих задач із використанням сучасних методів аналізу даних.	ПРН 04
Формулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, та обирати метод її розв'язання, що забезпечує потрібні точність і надійність результату.	ПРН 05
Уміти будувати комп'ютерний експеримент для конкретних задач прикладної математики шляхом використання спеціалізованих (у тому числі й створених) програмних засобів, та виконувати опис та аналіз результатів експерименту.	ПРН 06
Вміти спроектувати архітектуру системи з великими обсягами даних, моделювати штучні нейронні мережі та використовувати їх на практиці, застосовувати технологію блокчейн.	ПРН 10

Очікувані результати навчання, які повинні бути досягнуті після опанування навчальної дисципліни «Аналітика великих даних»:

Очікувані результати навчання з дисципліни	Шифр ПРН
Знати визначення великих даних та приклади великих даних у реальному світі	ПРН 10
Вміти експортувати та імпортувати дані в R та обробляти їх	ПРН 02, ПРН 04, ПРН 05, ПРН 06
Використовувати розподілену потокову платформу Kafka та знати призначення та переваги Cassandra	ПРН 01, ПРН 06
Застосовувати інструменти Spark для аналітичної обробки великих даних	ПРН 02, ПРН 04, ПРН 05
Знати переваги і недоліки Lambda -архітектури та Карра-архітектури.	ПРН 01, ПРН 10

5. ЗАСОБИ ДІАГНОСТИКИ ТА КРИТЕРІЇ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

Засоби оцінювання та методи демонстрування результатів навчання

Засобами оцінювання та методами демонстрування результатів навчання з навчальної дисципліни «Аналітика великих даних» є: виконання лабораторних робіт, модульні контрольні роботи, екзамен.

Форми (методи) контролю та критерії оцінювання результатів навчання

Форми поточного контролю: усні відповіді на лабораторних заняттях, захист лабораторних робіт.

Форма модульного контролю: письмова контрольна робота.

Форми підсумкового семестрового контролю: екзамен.

Розподіл балів, які отримують здобувачі вищої освіти (модуль 1)

Поточне оцінювання та самостійна робота					Модульна контрольна робота	Сума
T1	T2	T3	T4	T5	40	100
6	8	16	16	14		

T1, T2 ... – теми

Розподіл балів, які отримують здобувачі вищої освіти (модуль 2)

Поточне оцінювання та самостійна робота						Модульна контрольна робота	Сума
T1	T2	T3	T4	T5	T6	40	100
8	10	12	10	12	8		

T1, T2 ... – теми

Оцінювання окремих видів навчальної роботи з дисципліни

Вид діяльності здобувача вищої освіти	Модуль 1		Модуль 2	
	Кількість	Максимальна кількість балів (сумарна)	Кількість	Максимальна кількість балів (сумарна)
Лабораторні заняття (допуск, виконання та захист)	3	60	3	60
Модульна контрольна робота	1	40	1	40
Разом		100		100

Критерії оцінювання модульної контрольної роботи

Методика оцінювання. Матеріал кожного модуля, який здобувачі вищої освіти повинні засвоїти протягом семестру, виноситься на одну з двох модульних контрольних робіт.

Модульна контрольна робота складається із 4-ох завдань (2-ох теоретичних питань та 2-ох практичних завдань), кожне з яких оцінюється в 10 балів.

За виконання лабораторних робіт здобувачу вищої освіти також нараховується різна кількість балів, в залежності від складності матеріалу.

Критерієм успішного проходження здобувачем освіти поточного оцінювання (включно із захистом лабораторних робіт) є досягнення здобувачем освіти не менше 50% балів від загальної кількості запланованої за конкретною темою. Конкретна максимальна кількість балів подається у таблицях розподілу балів, які отримують здобувачі за модуль та за окремі види навчальної роботи.

Невиконані та незахищені лабораторні роботи, а також неявка на модульну контрольну роботу оцінюються в 0 балів незалежно від причини невиконання (неявки).

Сумарна оцінка (від 0 до 100 балів) виставляється у відомість модульного контролю. Модуль зараховується, якщо сумарний бал складає не менше 60 балів, і студент виконав і захистив всі лабораторні роботи, які є складовими даного модуля.

Здобувач вищої освіти, який не з'явився на модульну контрольну роботу, або ж його модульна оцінка складає від 0 до 34 балів, зобов'язаний скласти (перескласти) модуль до початку підсумкового контролю у строки, визначені викладачем дисципліни та погоджені деканатом факультету.

Критерії оцінювання підсумкового контролю

Екзаменаційна методика оцінювання. За результатами модульних контролів визначається підсумкова модульна оцінка. Екзаменаційна оцінка визначається в залежності від рейтингового балу, або балів за екзамен.

До складання екзамену допускаються здобувачі вищої освіти, у яких підсумкова модульна оцінка за семестр становить не менше 35.

Здобувач вищої освіти, підсумкова модульна оцінка якого складає від 0 до 34 балів, зобов'язаний покращити її до початку підсумкового семестрового контролю під час чергування викладача на кафедрі у строки, визначені викладачем дисципліни та погоджені деканатом факультету. В протилежному випадку, здобувач не допускається до екзамену і у нього виникає академічна заборгованість.

Екзамен з навчальної дисципліни здобувач вищої освіти може не скласти, якщо він успішно пройшов усі модульні контролі та його влаштовує підсумкова модульна оцінка. Здобувачі вищої освіти, підсумкова модульна оцінка яких становить від 35 до 59, екзамен складають обов'язково. Здобувач освіти може підвищити на екзамені рейтинговий бал, при цьому, за результатами складання екзамену оцінка не може бути менша за підсумкову модульну оцінку, яку він отримав за результатами модульних контролів.

Екзамен проводиться в письмовій формі. Екзаменаційний білет складається з двох теоретичних питань та двох практичних завдань. Оцінювання результатів навчання на екзамену здійснюється за 100-бальною шкалою. Оцінка за екзамен вноситься у відомість обліку успішності.

Таблиця відповідності оцінок за різними шкалами оцінювання

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену	для заліку
90-100	A	відмінно	зараховано
82-89	B	добре	
74-81	C		
64-73	D	задовільно	
60-63	E		
35-59	FX	незадовільно з можливістю повторного складання	незараховано з можливістю повторного складання
0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	незараховано з обов'язковим повторним вивченням дисципліни

Критерій оцінювання підсумкового контролю з дисципліни

— **«відмінно» (90-100 балів, A)** заслуговує студент, який виявив всебічне і глибоке знання програмового матеріалу, вміння вільно виконувати завдання, передбачені програмою, засвоїв основну і ознайомився з додатковою літературою, розуміє взаємозв'язок головних понять дисципліни та їх значення для майбутньої професії;

— **«добре» (82-89 балів, B)** заслуговує студент, який виявив повне знання програмного матеріалу, успішно виконує передбачені програмою завдання, засвоїв основну літературу рекомендовану програмою, виявив систематичний характер знань з дисциплін і здатний до самостійного доповнення, але під час відповіді допустив деякі неточності;

— **«добре» (74-81 бал, C)** заслугоує студент, що виявив не цілком повне знання програмного матеріалу, не завжди успішно виконує передбачені програмою завдання, частково засвоїв основну літературу, рекомендовану програмою, виявив не систематичний характер знань з дисциплін і не завжди здатний до їх самостійного доповнення і під час відповіді допускає деякі неточності;

— **«задовільно» (64-73 бали, D)** заслугоує студент, що виявив знання основного програмного матеріалу в обсязі, необхідному для подальшого навчання та майбутньої роботи за професією, вміє виконувати завдання, передбачені програмою, знайомий з основною рекомендованою літературою. Як правило, дана оцінка виставляється студентам, що допустили помилки у відповіді на заліку чи екзамені та при виконанні залікових або екзаменаційних завдань, але які володіють необхідними знаннями для їх усунення за допомогою викладача;

— **«задовільно» (60-63 балів, E)** заслугоує студент, що виявив часткове знання основного програмного матеріалу в обсязі, необхідному для подальшого навчання та майбутньої роботи за професією, не завжди вміє виконувати завдання, передбачені програмою, знайомий лише частково з основною рекомендованою літературою. Як правило, дана оцінка виставляється студентам, що допустили грубі помилки у відповіді на заліку чи екзамені та при виконанні залікових або екзаменаційних завдань, але які частково володіють необхідними знаннями для їх усунення за допомогою викладача.

— **«незадовільно» (35-59 балів, FX)** виставляється студенту, який виявив суттєві прогалини в знаннях основного програмного матеріалу, допустив принципові помилки у виконанні передбачених програмою завдань.

— **«незадовільно» (0-34 балів, F)** виставляється студенту коли протягом семестру він допустив грубі помилки у виконанні передбачених програмою завдань.

6. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

6.1. Зміст навчальної дисципліни

Модуль 1.

Тема 1. Джерела великих даних. Інтернет Речей. Визначення Big Data.

Інтернет Речей та зростання даних. Платформа Kaggle. DrivenData. Визначення великих даних. Приклади великих даних у реальному світі. Відкриті дані. Приватність даних. Структуровані та неструктуровані дані. Хмарні та туманні обчислення. Дані в спокої та дані в русі. Інфраструктура великих даних. Розподілені дані та їх обробка.

Тема 2. Основи роботи з середовищем програмування R.

Тема 3. Типи даних та елементарні функції. Цикли, умови, створення власних функцій.

Тема 4. Експорт, імпорт та оброблення даних в R.

Тема 5. Основні інструменти аналізу та візуалізації даних в R.

Модуль 2.

Тема 1. Архітектурні моделі Big Data. Технології віртуалізації. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері.

Архітектурні моделі інженерії Big Data. Центри обробки даних та хмарні обчислення. Технології віртуалізації. Шари абстракції. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері. Інжиніринг даних.

Тема 2. Технології Hadoop Big Data. Розподілена обробка MapReduce.

Масштабованість за допомогою великих даних. Зберігання та оброблення даних в розподілених файлових системах. Розподілені бази даних. Розподілена файлова система Hadoop.

Тема 3. Розподілена потокова платформа Kafka. Переваги Cassandra.

Тема 4. Платформа Apache Spark.

Проблема обчислювальної функції. Технологія Spark. Порівняння Spark та MapReduce.

Тема 5. Spark і sparklyt для роботи з великими даними в R.

Тема 6. Lambda та Карра архітектури оброблення великих даних.

Lambda - архітектура. Переваги і недоліки Lambda -архітектури. Карра - архітектура. Переваги і недоліки Карра-архітектури.

6.2. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин					
	Денна форма					
	Усього	у тому числі				
		лекції	практичні	лабораторні	індивідуальна робота	самостійна робота
1 семестр						
Модуль 1						
Тема 1. Джерела великих даних. Інтернет Речей. Визначення Big Data.	9	4				5
Тема 2. Основи роботи з середовищем програмування R.	10	2				8
Тема 3. Типи даних та елементарні функції. Цикли, умови, створення власних функцій.	14	2		4		8
Тема 4. Експорт, імпорт та оброблення даних в R.	14	2		4		8
Тема 5. Основні інструменти аналізу та візуалізації даних в R.	13	2		4		7
Разом за модуль	60	12		12		36
Модуль 2						
Тема 1. Архітектурні моделі Big Data. Технології віртуалізації. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері.	7	2				5
Тема 2. Технології Hadoop Big Data. Розподілена обробка MapReduce.	11	2		4		5
Тема 3. Розподілена потокова платформа Kafka. Переваги Cassandra.	12	2		4		6
Тема 4. Платформа Apache Spark.	8	2				6
Тема 5. Spark і sparklyt для роботи з великими даними в R.	12	2		4		6
Тема 6. Lambda та Карра архітектури оброблення великих даних.	10	2				8
Разом за модуль	60	12		12		36
Разом за семестр	120	24		24		72

6.3. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1	Типи даних та елементарні функції. Цикли, умови, створення власних функцій.	4

2	Експорт, імпорт та оброблення даних в R.	4
3	Основні інструменти аналізу та візуалізації даних в R.	4
4	Технології Hadoop Big Data. Розподілена обробка MapReduce.	4
5	Розподілена потокова платформа Kafka. Переваги Cassandra.	4
6	Spark і sparklyr для роботи з великими даними в R.	4
Разом		24

6.4. Самостійна робота

№ з/п	Назва теми	Кількість годин
1	Джерела великих даних. Інтернет Речей. Визначення Big Data.	5
2	Основи роботи з середовищем програмування R.	8
3	Типи даних та елементарні функції. Цикли, умови, створення власних функцій.	8
4	Експорт, імпорт та оброблення даних в R.	8
5	Основні інструменти аналізу та візуалізації даних в R.	7
6	Архітектурні моделі Big Data. Технології віртуалізації. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері.	5
7	Технології Hadoop Big Data. Розподілена обробка MapReduce.	5
8	Розподілена потокова платформа Kafka. Переваги Cassandra.	6
9	Платформа Apache Spark.	6
10	Spark і sparklyr для роботи з великими даними в R.	6
11	Lambda та Карра архітектури оброблення великих даних.	8
Разом		72

7. ІНСТРУМЕНТИ, ОБЛАДНАННЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ, ВИКОРИСТАННЯ ЯКИХ ПЕРЕДБАЧАЄ НАВЧАЛЬНА ДИСЦИПЛІНА

Технічні засоби – персональні комп'ютери, мультимедійний проектор, інтерактивна дошка.

Програмне забезпечення – операційна система; програмне середовище для статистичних обчислень, аналізу та зображення даних в графічному вигляді R; вільне та відкрите інтегроване середовище розробки RStudio; платформа Hadoop для програмування з відкритим кодом на основі Java, яка підтримує зберігання та обробку наборів великих даних у розподіленому обчислювальному середовищі; сервіс Google Meet; система електронного навчання Moodle.

8. РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

Основна література

1. Кудін О.В. Моделювання систем та аналіз даних: методичні рекомендації до лабораторних робіт для студентів освітнього ступеня «бакалавр» напряму підготовки «Програмна інженерія». Запоріжжя: ЗНУ, 2017. 89 с.

2. Майборода Р. Є. Комп'ютерна статистика : підручник. К.:ВПЦ "Київський університет", 2019. 589 с.
3. 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. Київ: КПІ ім. Ігоря Сікорського, 2021. 227 с.
4. Benjamin S. Baumer, Daniel T. Kaplan , Nicholas J. Horton. Modern Data Science with R. 2nd Edition. Chapman and Hall/CRC, 2021. 632 pp.
5. Hadley Wickham, Mine Çetinkaya-Rundel, Garrett Grolemund. R for Data Science: Import, Tidy, Transform, Visualize, and Model. O'Reilly Media, Inc., 2023. 576 pp.

Допоміжна література

1. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: <https://www.netacad.com/courses/iot/big-data-analytics>
2. IoT Analytics Platform // Електронний ресурс. Режим доступу: <https://blog.codecentric.de/en/2016/07/iot-analytics-platform/>
3. Akka // Електронний ресурс. Режим доступу: <https://akka.io/>
4. IoT-Analyse-Plattform: Floating Bus Data // Електронний ресурс. Режим доступу: https://www.youtube.com/watch?v=VYxc-3ZRRL4&ab_channel=codecentricAG
5. A brief introduction to two data processing architectures — Lambda and Kappa for Big Data // Електронний ресурс. Режим доступу: <https://towardsdatascience.com/a-briefintroduction-to-two-data-processing-architectures-lambda-and-kappa-for-big-data-4f35c28005bb>