МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ І.І. МЕЧНИКОВА
Кафедра математичного забезпечення комп'ютерних систем

"ЗАТВЕРДЖУЮ"
Проректор з науково-педагогічної роботи
_____
"___" _____ 20___ р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ОК10 «Аналіз та візуалізація надвеликих масивів даних (Big Data)»
(назва навчальної дисципліни)

Рівень вищої освіти *Другий (магістерський)*

Галузь знань *12 – Інформаційні технології*

Спеціальність *126 – Інформаційні системи та технології*
(код і назва спеціальності)

Освітньо-професійна програма *Інформаційні системи та технології*
(назва ОПП/ОНП)

ОНУ – 2023

Робоча програма навчальної дисципліни «Аналіз та візуалізація надвеликих масивів даних (Big Data)». – Одеса: ОНУ, 2023. – 9 с.

Розробник:
Петрушина Т.І., к.ф.-м.н., доцент кафедри МЗКС

Робоча програма затверджена на засіданні кафедри математичного забезпечення комп'ютерних систем

Протокол № 4 від. "7" 11 2023 р.

Завідувач кафедри _____ (підпис) (Євгеній МАЛАХОВ) (Ім'я ПРІЗВИЩЕ)

Погоджено із гарантом ОПП «Інформаційні системи та технології»

_____ (підпис) (Євгеній МАЛАХОВ) (Ім'я ПРІЗВИЩЕ)

Схвалено навчально-методичною комісією (НМК) з ІТ спеціальностей факультету МФІТ

Протокол № 3 від. "7" 11 2023р.

Голова НМК _____ (підпис) (Алла РАЧИНСЬКА) (Ім'я ПРІЗВИЩЕ)

Переглянуто та затверджено на засіданні кафедри _____

Протокол № ___ від. "____"_____20__ р.

Завідувач кафедри _____ (підпис) (_____) (Ім'я ПРІЗВИЩЕ)

Переглянуто та затверджено на засіданні кафедри _____

Протокол № ___ від. "____"_____20__ р.

Завідувач кафедри _____ (підпис) (_____) (Ім'я ПРІЗВИЩЕ)

1. Опис навчальної дисципліни

<table>
<tr><td rowspan="2">Найменування показників</td><td rowspan="2">Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень</td><td colspan="2">Характеристика навчальної дисципліни</td></tr>
<tr><td>денна форма навчання</td><td>заочна форма навчання</td></tr>
<tr><td rowspan="14">Загальна кількість:

кредитів – 4

годин – 120

змістових модулів – 3</td><td rowspan="14">Галузь знань
<u>12 – Інформаційні технології</u>

Спеціальність
<u>126 – Інформаційні системи та технології</u>

Рівень вищої освіти:
<u>Другий (магістерський)</u></td><td colspan="2">Обов’язкова</td></tr>
<tr><td colspan="2">Рік підготовки:</td></tr>
<tr><td>1-й</td><td></td></tr>
<tr><td colspan="2">Семестр</td></tr>
<tr><td>2-й</td><td></td></tr>
<tr><td colspan="2">Лекції</td></tr>
<tr><td>18 год.</td><td>8 год.</td></tr>
<tr><td colspan="2">Практичні, семінарські</td></tr>
<tr><td>год.</td><td>год.</td></tr>
<tr><td colspan="2">Лабораторні</td></tr>
<tr><td>18 год.</td><td>6 год.</td></tr>
<tr><td colspan="2">Самостійна робота</td></tr>
<tr><td>84 год.</td><td>106 год.</td></tr>
<tr><td colspan="2">Індивідуальні завдання: год.</td></tr>
<tr><td colspan="2">Форма підсумкового контролю: іспит</td></tr>
</table>

* – за наявності

2. Мета та завдання навчальної дисципліни

Метою викладання курсу є формування у студентів представлення про технічні та методологічні засоби організації, зберігання та аналізу надвеликих даних (Big Data).

Завдання: набуття компетентностей на основі засвоєння основних теоретичних положень щодо існуючих технічних та методологічних засобів аналізу надвеликих даних, що забезпечують зберігання і управління обсягом даних в сотні терабайт або петабайт, які звичайні реляційні бази даних не дозволяють ефективно використовувати; освоєння ефективних методів організації неструктурованою інформації (текст, зображення, відео та ін), засобів і способів роботи з нею, генерацію аналітичних звітів, впровадження прогностичних моделей.

Процес вивчення дисципліни спрямований на формування елементів наступних **компетентностей**:

а) загальних:

—

б) фахових:

СК04. Здатність розробляти математичні, інформаційні та комп'ютерні моделі об'єктів і процесів інформатизації.

СК05. Здатність використовувати сучасні технології аналізу даних для оптимізації процесів в інформаційних системах.

СКМ07. Здатність проводити інформаційний аналіз та створювати багатовимірні моделі предметних областей.

Програмні результати навчання:

РН09. Розробляти і використовувати сховища даних, здійснювати аналіз даних для підтримки прийняття рішень.

РН11. Розв'язувати задачі цифрової трансформації у нових або невідомих середовищах на основі спеціалізованих концептуальних знань, що включають сучасні наукові здобутки у сфері інформаційних технологій, досліджень та інтеграції знань з різних галузей.

РНМ07. Розробляти та забезпечувати підтримку автономних розподілених інтелектуальних системи автоматизованого пошуку та аналізу інформації.

РНМ08. Створювати оптимізовані пайплайни для підготовки даних до подальшого зберігання та обробки.

Очікувані результати навчання. У результаті вивчення навчальної дисципліни студент повинен

***знати**: основні поняття, методи та засоби в області Big Data; основи машинного навчання, візуалізації та зберігання великих даних; основи роботи з хранилищами даних і NoSQL СУБД*

***вміти**: застосовувати підходи аналітичної обробки великих даних при вирішенні задач, пов'язаних з управлінням у складних технічних системах;*

освоювати методики використання програмних засобів, що підтримують технології Big Data для вирішення практичних задач в предметній області; перевести проблеми предметної області на технологію обробки великих даних.

3. Зміст навчальної дисципліни

Змістовний модуль 1. Загальні поняття та визначення Big Data.

Тема 1. Сучасні підходи до обробки та зберігання надвеликих даних.

Тема 2. Передобробка даних, візуалізація, первинний статистичний аналіз.

Тема 3. Форми представлення даних, типи і види даних.

Змістовний модуль 2. Програмне забезпечення в області аналізу великих даних.

Тема 1. Аналітичні платформи: класифікація та особливості застосування.

Тема 2. Мови візуального моделювання. Мова R

Тема 3. NoSQL СУБД. Hadoop.

Тема 4. MapReduce. Повнотекстовий пошук. ETL процес по обробці звітів.

Змістовний модуль 3. Технології KDD і Data Mining.

Тема 1. Підготовка даних до аналізу. Методика вилучення знань.

Тема 2. Пошук асоціативних правил. Застосування класифікації і регресії. Статистичні методи.

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин									
	Денна форма					Заочна форма				
	Усього	у тому числі				Усього	у тому числі			
		л	п	лаб	ср		л	п	лаб	ср
1	2	3	4	5	6	7	8	9	10	11
Змістовний модуль 1. Загальні поняття та визначення Big Data.										
Тема 1.	5	1			4	6				6
Тема 2.	5	1			4	10	1		1	8
Тема 3.	5	1			4	10				8
Змістовний модуль 2. Програмне забезпечення в області аналізу великих даних										
Тема 1.	10	2			8	13,5	1		0,5	12
Тема 2.	14	2		2	10	13			1	12
Тема 3.	15	3		2	10	15,5	1		0,5	14
Тема 4.	19	3		4	12	18	1		1	16
Змістовний модуль 3. Технології KDD і Data Mining										
Тема 1.	23	3		4	16	16	1		1	14
Тема 2.	24	2		6	16	18	1		1	16
Всього годин	**120**	**18**		**18**	**84**	**120**	**8**		**6**	**106**

5. Теми семінарських занять

Семінарські заняття не передбачені

6. Теми практичних занять

Практичні заняття не передбачені

7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин	
		Денна форма	Заочна форма
1	Підготовка, обробка, аналіз та візуалізація результатів з допомогою MS SQL Analytical Services.	4	2
2	Кластерний аналіз у середовищі R	4	2
3	Регресійний аналіз в середовищі R	6	2
4	Байєсова класифікація в середовищі R	4	
	Разом	**18**	**6**

8. Самостійна робота

№ з/п	Назва теми / види завдань	Кількість годин	
		Денна форма	Заочна форма
1	Аналіз існуючих програмних інструментів Big Data.[1]	10	12
2	Специфіка застосування технології KDD до великих даних. [1]	19	25
3	Застосування нейронних мереж до аналізу великих даних.[1]	19	25
4	Організація обробки даних в HDFS.[1]	21	29
5	РГР – Індивідуальне завдання з проведення аналізу даних в середовище R	15	15
	Разом	**84**	**106**

До самостійної роботи відноситься:
[1] – підготовка до лекцій та лабораторних занять;

8.1. Індивідуальне навчально-дослідне завдання (курсовий проект або розрахунково-графічна робота) не передбачено

9. Методи навчання

Лекції з використанням мультимедійного презентаційного матеріалу.

10. Методи контролю

Методи **поточного контролю**: оцінювання виконання лабораторних робіт.

Підсумковий контроль: Іспит. Під час підсумкового контролю студент повинен відповісти на теоретичні питання з переліку, наведеному у п. 11.

10.1. Критерії оцінювання на підсумковому модульному контролі:

1. Відповідь повинна бути повною і короткою. Вона не повинна мати в собі матеріал, що не відноситься до сутті питання.
2. Чітко формулювати твердження, вправно застосовувати необхідні формули і знання основних питань програми.
3. Відповіді, що мають помилкові твердження оцінюються виходячи з близькості відповіді до правильної.
4. Пропуски в обґрунтуванні тверджень враховуються і це призводить до зменшення кількості балів.
5. Малі недоліки, неточності при викладенні матеріалу, зменшують кількість балів.
6. Незнання і нерозуміння основної ідеї теоретичного питання або задачі призводить до зняття до 90 % балів.
7. Якщо відповідь на питання відсутня то виставляється нуль балів.

11. Питання до підсумкового контролю

1. Визначте сутність поняття «великі дані».
2. Охарактеризуйте методики аналізу великих даних.
3. Опишіть процес аналізу великих даних.
4. Які особливості зберігання великих даних.
5. Дайте характеристики Big Data на світовому ринку.
6. Визначте поняття Data Mining.
7. Визначте поняття KDD
8. Визначте відмінності між параметричними, непараметричними і номінальними методами.
9. Які завдання підтримки безпеки великих даних.
10. У чому полягає когнітивний аналіз даних.
11. Які моделі даних ви знаєте?
12. Основні описові статистики.
13. Особливості застосування кореляційно-регресивного аналізу великих даних.
14. Сутність кластерного аналізу в застосуванні до великих даними.
15. Пошук асоціативних правил в великих даних.
16. Класифікація даних за допомогою нейромережі.
17. Класифікація за допомогою дерев рішень.
18. Програмні засоби аналізу великих даних і їх недоліки.
19. Основні можливості зберігання великих даних на мові програмування R.
20. Основні можливості мови R для аналізу великих даних.

12. Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота									Підсумковий контроль (іспит)	Сума
Змістовий модуль 1			Змістовий модуль 2			Змістовий модуль 3				
Т1	Т2	Т3	Т1	Т2	Т3	Т4	Т1	Т2		
8	8	8	8	10	10	8	10	10	20	100

Т1, Т2 ... – теми змістових модулів

Шкала оцінювання: національна та ECTS

Загальна сума балів	Оцінка ECTS	Національна шкала	
90 — 100	A – «відмінно»	5 «відмінно»	«залік»
85 — 89	B – «дуже добре»	4 «добре»	
75 — 84	C – «добре»		
70 — 74	D – «задовільно»	3 «задовільно»	
60 — 69	E – «допустимо»		
35 — 59	F – «незадовільно з можливістю повторного складання»	2 «незадовільно»	«незалік»
0 — 34	FX – «незадовільно з обов’язковим повторним курсом»		

13. Методичне забезпечення

Конспект лекцій в електронному форматі; комплекс навчально-методичного забезпечення дисципліни; нормативні документи; презентаційні матеріали.

14. Рекомендована література

14.1. Основна література

1. Weigend A. Data for the People. – Basic Books, 2017. – 357 p.
2. Cielen D., Introducing Data Science: Big Data, Machine Learning, and more, using Python tools // D. Cielen, A. Meysman, M. Ali / Manning, 2016. — 320 p.
3. Coelho L.P. Building Machine Learning Systems with Python / L.P. Coelho, W. Richert, M. Brucher // 3rd edition. – Packt Publishing, 2018, – 408 p.
4. Інтелектуальний аналіз даних: Підручник / Черняк О.І., Захарченко П.В./ К.: Знання, 2014р. - 599 с.
5. Акіменко В.В., Загородній Ю.В. Проектування СППР на основі нечіткої логіки. Навчально- методичний посібник. – К.: Вид-во КНУ, 2007. – 94с.

14.2. Допоміжна література

1. Майер-Шенбергер, В.Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер ; пер. с англ. Инны Гайдюк. — М. : Манн, Иванов и Фербер, 2014. — 240 с.
2. Уайт Т. Hadoop: Подробное руководство. — СПб.: Питер, 2013. — 672 с.:ил.
3. Чубукова И. А. Data Mining: учебное пособие. — М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. — 382 с
4. Чак Л. Hadoop в действии. - М.: ДМК Пресс, 2012. - 424с: ил.
5. Шипунов А.Б., Балдин Е.М., Волкова П.А., Коробейников А.И., Назарова С.А., Петров С.В., Суфиянов В.Г. Наглядная статистика. Используем R! - М.: ДМК Пресс, 2012. - 298с.

15. Електронні інформаційні ресурси

1. Великі дані. [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/Великі_дані
2. Big Data for Development: From Information- to Knowledge Societies [Електронний ресурс]. – Режим доступу: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205145
3. Big Data от А до Я. [Електронний ресурс]. – Режим доступу: https://habrahabr.ru/company/dca/blog/267361/
4. Big Data и блокчейн. [Електронний ресурс]. – Режим доступу:
5. https://forklog.com/big-data-i-blokchejn-proryv-v-oblasti-analiza-dannyh/
6. Онлайн курс Machine learning https://www.coursera.org/course/ml
7. Онлайн курс Big Data Overview https://education.emc.com/academicalliance/elearning/Big_Data_Overview/index.htm
8. Онлайн курс R programming https://www.coursera.org/course/rprog
9. Онлайн курс Introduction to Data Science https://www.coursera.org/course/datasci
10. Онлайн курс «Введение в аналитику больших массивов данных» http://bit.ly/IntuitBDA.
11. Учебник по статистическому обучению http://statweb.stanford.edu/~tibs/ElemStatLearn/