

Міністерство освіти і науки України  
Харківський національний університет імені В.Н. Каразіна  
Кафедра прикладної математики

“ЗАТВЕРДЖУЮ”

Декан факультету  
математики і інформатики

Григорій ЖОЛТКЕВИЧ



” 08 2023 р.

Робоча програма навчальної дисципліни

**Прикладні задачі аналізу "великих даних"**

(назва навчальної дисципліни)

рівень вищої освіти бакалавр

галузь знань 11 – Математика та статистика

спеціальність 113 – Прикладна математика  
(шифр і назва)

освітня програма Прикладна математика  
(шифр і назва)

спеціалізація \_\_\_\_\_  
(шифр і назва)

вид дисципліни за вибором  
(обов'язкова / за вибором)

факультет математики і інформатики

2023 / 2024 навчальний рік

Кафедра прикладної математики

Програму рекомендовано до затвердження вченою радою факультету математики і інформатики  
 “29” серпня 2023 року, протокол №8

РОЗРОБНИКИ ПРОГРАМИ: (вказати авторів, їхні наукові ступені, вчені звання та посади)  
**Кізілова Наталія Миколаївна**, доктор фіз.-мат. наук, професор кафедри прикладної математики.

Програму схвалено на засіданні кафедри Прикладної математики  
 Протокол від “28” серпня 2023 року №10

Завідувач кафедри Прикладної математики



(підпис)

Валерій КОРОБОВ

(прізвище та ініціали)

Програму погоджено з гарантом освітньо-професійної програми (керівником проектної групи) Прикладна математика  
 назва освітньої програми

Гарант освітньо-професійної програми  
 (керівник проектної групи) Прикладна математика



(підпис)

Сергій ПОСЛАВСЬКИЙ

(прізвище та ініціали)

Програму погоджено науково-методичною комісією факультету математики і інформатики Харківського національного університету імені В.Н. Каразіна  
 назва факультету, для здобувачів вищої освіти якого викладається навчальна дисципліна

Протокол від “29” серпня 2023 року, протокол №1

Голова науково-методичної комісії факультету математики і інформатики



(підпис)

Ольга АНОЩЕНКО

(прізвище та ініціали)

## ВСТУП

Програма навчальної дисципліни “**Прикладні задачі аналізу великих даних**” складена відповідно до освітньо-наукової програми підготовки  
 \_\_\_\_\_  
 бакалавр  
 \_\_\_\_\_  
 (назва рівня вищої освіти)

спеціальності \_\_\_\_\_ 113 Прикладна математика \_\_\_\_\_

спеціалізації \_\_\_\_\_

### 1. Опис навчальної дисципліни

#### 1.1. Мета викладання навчальної дисципліни

Метою викладання навчальної дисципліни “**Прикладні задачі аналізу великих даних**” є надання знань в галузі сучасного статистичного аналізу «великих даних», відповідних математичних і аналітичних моделей, їх програмної реалізації та практичних застосувань у рішенні конкретних прикладних задач.

#### 1.2. Основні завдання вивчення дисципліни.

Основними завданнями вивчення дисципліни “**Прикладні задачі аналізу великих даних**” є вивчення студентами основних статистичних методів, математичних моделей та підходів у аналізі різних типів «великих даних», знання прикладних пакетів програм для рішення актуальних прикладних проблем.

1.3. Кількість кредитів – 12

1.4. Загальна кількість годин – 120

1.5. Характеристика навчальної дисципліни	
За вибором	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
4-й	-й
Семестр	
8-й	-й
Лекції	
32 год	Год
Практичні, семінарські заняття	
22 год	Год
Лабораторні заняття	
	Год
Самостійна робота	
66	Год
у тому числі індивідуальні завдання	
	Год

## 1.6. Заплановані результати навчання

студенти повинні досягти таких результатів навчання:

### ***Знати:***

1. основні методи та підходи до аналізу «великих даних»;
2. статистичні та аналітичні методи досліджень «великих даних»;
3. сучасні пакети прикладних програм для статистичного аналізу «великих даних»;
4. структуру та особливості роботи з базами «великих даних» в мережі Інтернет;

### ***Вміти:***

1. використовувати сучасні статистичні методи для аналізу «великих даних» в різних прикладних задачах;
2. вміти користуватися існуючими програмними пакетами та розробляти власний програмний код для чисельної реалізації аналізу «великих даних»;
3. розробляти системи штучного інтелекту для аналізу «великих даних»;
4. проводити пошук інформації у вигляді «великих даних» в мережі Інтернет.

## **2. Тематичний план навчальної дисципліни.**

### ***Розділ 1. Статистичні методи та аналітичні моделі аналізу «великих даних».***

**Тема 1.** Види і структура різних типів «великих даних».

**Тема 2.** Основні статистичні методи аналізу «великих даних».

**Тема 3.** Головні аналітичні моделі науки о даних.

**Тема 4.** Методи та алгоритми штучного інтелекту в аналізі «великих даних».

### ***Розділ 2. Конкретні прикладні задачі обробки та аналізу «великих даних».***

**Тема 5.** Методи «великих даних» для аналізу медичної інформації.

**Тема 6.** Реалізація аналізу «великих даних» в біоінформатиці.

**Тема 7.** Реалізація аналізу «великих даних» в геоінформатиці.

**Тема 8.** Реалізація методів «великих даних» при аналізі економічної інформації.

## 3. Структура навчальної дисципліни

Назви розділів і тем	Кількість годин											
	Денна форма						Заочна форма					
	Усього	у тому числі					Усього	у тому числі				
		л	п	лаб	інд	ср		л	п	лаб	інд	ср
1	2	3	4	5	6	7	8	9	10	11	12	13
<b>Розділ 1. Статистичні методи та аналітичні моделі аналізу «великих даних»</b>												
Тема 1. Види і структура різних типів «великих даних»	16	4	3			9						
Тема 2. Основні статистичні методи аналізу «великих даних»	15	4	3			8						
Тема 3. Головні аналітичні моделі науки о даних	15	4	3			8						
Тема 4. Методи та алгоритми штучного інтелекту в аналізі «великих даних»	14	4	2			8						
Разом за розділом 1	60	16	11			33						
<b>Розділ 2. Конкретні прикладні задачі обробки та аналізу «великих даних».</b>												
Тема 5. Методи «великих даних» для аналізу медичної інформації	16	4	3			9						
Тема 6. Реалізація аналізу «великих даних» в біоінформатиці	15	4	3			8						
Тема 7. Реалізація аналізу «великих даних» в геоінформатиці	15	4	3			8						
Тема 8. Реалізація методів «великих даних» при аналізі економічної інформації	14	4	2			8						
Разом за розділом 2	60	16	11			33						
<b>Всього годин</b>	<b>120</b>	<b>32</b>	<b>22</b>			<b>66</b>						

#### 4. Теми семінарських (практичних, лабораторних) занять

№ з/п	Назва теми	Кількість годин
1	Структура масивів «великих даних» на прикладах	2
2	Регресивний, кореляційний, факторний і кластерний аналіз даних.	2
3	Методи аналізу та розпізнавання зображень	2
4	Аналітичні моделі для дослідження «великих даних»	2
5	Алгоритми глибокого навчання і штучного інтелекту в аналізі «великих даних»	4
6	<i>Контрольна робота</i>	2
7	Типи медичної інформації та засоби її обробки	4
8	Типи біологічної інформації та засоби її обробки	2
9	Типи геофізичної інформації та засоби її обробки	2
	<b>Разом</b>	<b>22</b>

#### 5. Завдання для самостійної роботи

№ з/п	Види, зміст самостійної роботи	Кількість годин
1	Виконання домашніх завдань за темою «Інтернет речей (Internet of Things) та методи аналізу «великих даних»»	20
2	Написання та валідація програмного коду для аналізу конкретної медичної інформації	22
3	Написання та валідація програмного коду для аналізу медичних зображень	24
	<b>Разом</b>	<b>66</b>

#### 6. Індивідуальні завдання

*Не передбачено планом*

#### 7. Методи навчання

Лекційні та практичні заняття проводяться дистанційно за допомогою платформ Google Meet або Zoom.

#### 8. Методи контролю

- Перевірка виконання домашніх завдань
- Контрольна робота
- Проведення заліку.

#### Критерії оцінювання

**Контрольна робота** оцінюється у 20 балів. Робота складається з теоретичних та практичних запитань. У разі правильної обґрунтованої відповіді студент отримує за завдання бали; якщо у відповіді є помилки, бал не зараховується.

**Зміст семестрового проекту:** студентам надано файл даних з відкритих джерел у вигляді часових рядів для статистичного аналізу даних вказаним статистичним методом. Потрібно провести статистичний аналіз з використанням доступного або власно розробленого програмного коду, проаналізувати результати обробки і запропонувати відповідну математичну модель дослідженої системи.

Після виконання семестрового проекту студенти надають звіт, за результатами цього звіту виставляється оцінка до 40 балів за такими критеріями:

0-10 балів - Було проведено аналіз файлу даних, однак практичної реалізації не отримано.

10-25 балів - Було проведено аналіз файлу даних та запропоновано математичну модель задачі. Був використаний доступний метод обробки даних, але без відповідного аналізу даних.

25-35 балів - Було проведено аналіз файлу даних, статистичний аналіз даних з використанням власно розробленого програмного коду, який був протестований на відомому масиві даних.

35-40 балів - Було проведено аналіз файлу даних, статистичний аналіз даних з використанням власно розробленого програмного коду на базі чого була запропонована математична модель у вигляді системи диференціальних рівнянь.

**Екзаменаційна робота** складається з трьох завдань і передбачає письмову відповідь на два теоретичних питання зі списку, який надається студентам заздалегідь, а також розв'язання практичної задачі.

Кожне теоретичне завдання оцінюється максимально 15 балами, задача – 10 балами.

По кожному завданню залікової роботи нараховується:

- максимальний бал у разі правильно обґрунтованої відповіді;
- за незначні помилки оцінка зменшується від 10 до 30 відсотків;
- за значні логічні помилки оцінка зменшується до 50 відсотків, якщо хід міркувань в цілому правильний;
- якщо відповідь не відповідає жодному з критеріїв – виставляється 0 балів.

#### **Шкала оцінювання: чотирирівнева**

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка
90 – 100	відмінно
70-89	добре
50-69	задовільно
1-49	незадовільно

## **10. Рекомендована література**

### **Основна література**

1. Кізілова Н.М. Методи аналізу «великих даних». Методичні рекомендації з курсу «Прикладні задачі аналізу великих даних» для студентів спеціальності «Прикладна математика». Харків: ХНУ ім. В.Н. Каразіна. 2021. – 96 с.
2. Шаховська Н.Б., Болюбаш Ю.Я., Верес О.М. Організація великих даних у розподіленому середовищі. // Обчислювальна техніка та автоматизація. Вісник ДонНТУ. 2014. – № 2(27). - С.147–155.

### **Допоміжна література**

1. Lowman M. A Practical Guide to Analytics for Governments: Using Big Data for Good. John Wiley & Sons. 2017. 224p.
2. Prasad Y.L. Big Data Analytics. Made Easy. Notial Press. 2016. 228p.

## 11. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

1. Omar H., Jumaа A.K. Big Data Analysis Using Apache Spark MLlib and Hadoop HDFS with Scala and Java  
[https://www.researchgate.net/publication/332936006\\_Big\\_Data\\_Analysis\\_Using\\_Apache\\_Spark\\_MLlib\\_and\\_Hadoop\\_HDFS\\_with\\_Scala\\_and\\_Java/link/5ce5c3c7299bf14d95b1cfc0/download](https://www.researchgate.net/publication/332936006_Big_Data_Analysis_Using_Apache_Spark_MLlib_and_Hadoop_HDFS_with_Scala_and_Java/link/5ce5c3c7299bf14d95b1cfc0/download)
2. Marr B. Big Data in Practice  
[https://books.google.pl/books?id=UIU7jgEACAAJ&printsec=frontcover&hl=ru&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.pl/books?id=UIU7jgEACAAJ&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
3. Багато підручників з математичної статистики <http://statosphere.ru/books-arch/multistat/69-k-muller.html>