Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра прикладної математики

"ЗАТВЕРДЖУЮ"

Проректор з
науково-педагогічної роботи
Антон ПАНТЕЛЕЙМОНОВ

"__" березня 2020 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

2020 / 2021 навчальний рік

Програму рекомендовано до затвердження Вченою радою факультету математики і інформатики

"31" серпня 2020 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 "31" серпня 2020 року №12

Завідувач кафедри Прикладної математики

(підпис) Валерій КОРОБОВ (прізвище та ініціали)

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

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

(підпис) Світлана ІГНАТОВИЧ (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 "31" серпня 2020 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 аналізу «великих даних», відповідних математичних і аналітичних моделей, їх програмної реалізації та практичних застосувань у рішенні 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 видів «великих даних», знання прикладних пакетів програм для рішення конкретних прикладних проблем.

1.3. Кількість кредитів – 4

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	
32 год	Год
Лабораторні заняття	
	Год
Самостійна робота	
56	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 різні 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
Розділ 1. Статистичні методи та аналітичні моделі аналізу «великих даних»												
Тема 1. Види і структура різних типів «великих даних»	15	4	4			7						
Тема 2. Основні статистичні методи аналізу «великих даних»	15	4	4			7						
Тема 3. Головні аналітичні моделі науки о даних	15	4	4			7						
Тема 4. Методи та алгоритми штучного інтелекту в аналізі «великих даних»	15	4	4			7						
Разом за розділом 1	60	16	16			28						
Розділ 2. Конкретні прикладні задачі обробки та аналізу «великих даних».												
Тема 5. Методи «великих даних» для аналізу медичної інформації	15	4	4			7						
Тема 6. Реалізація аналізу «великих даних» в біоінформатиці	15	4	4			7						
Тема 7. Реалізація аналізу «великих даних» в геоінформатиці	15	4	4			7						
Тема 8. Реалізація методів «великих даних» при аналізі економічної інформації	15	4	4			7						
Разом за розділом 2	60	16	16			28						
Всього годин	**120**	**32**	**32**			**56**						

4. Теми семінарських (практичних, лабораторних) занять

№ з/п	Назва теми	Кількість годин
1	Структура масивів «великих даних» на прикладах	3
2	Регресивний, кореляційний, факторний і кластерний аналіз даних.	3
3	Методи аналізу та розпізнавання зображень	3
4	Аналітичні моделі для дослідження «великих даних»	3
5	Алгоритми алгоритми штучного інтелекту в аналізі «великих даних»	4
6	*Контрольна робота*	4
7	Типи медичної інформації та засоби ії обробки	4
8	Типи біологічної інформації та засоби ії обробки	4
9	Типи геофізичної інформації та засоби ії обробки	4
	Разом	32

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»»	18
2	Написання та валідація програмного коду для аналізу конкретної медичної інформації	18
3	Написання та валідація програмного коду для аналізу медичних зображень	20
	Разом	**56**

6. Індивідуальні завдання

7. Методи навчання

Лекційні та практичні заняття проводяться аудиторно. А у разі оголошення карантину заняття проводяться відповідно до Наказу ректора Харківського національного університету імені В.Н. Каразіна (аудиторно або дистанційно за допомогою платформ Google Meet або Zoom).

8. Методи контролю

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

9. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання					Залік	Сума
Розділ 1	Розділ 2	Контрольна робота, передбачена навчальним планом	Індивідуальне завдання	Разом		
Т1–Т4	Т5–Т8					
25	25	10		60	40	100

Т1, Т2 … – теми розділів.

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка	
	для чотирирівневої шкали оцінювання	для дворівневої шкали оцінювання
90 – 100	відмінно	зараховано
70-89	добре	
50-69	задовільно	
1-49	незадовільно	не зараховано

Критерії оцінювання навчальних досягнень

Оцінка в балах	Оцінка за національною шкалою	
Оцінка	Пояснення	
90 – 100	Відмінно	Теоретичний зміст курсу освоєний цілком, необхідні практичні навички роботи з освоєним матеріалом сформовані, всі навчальні завдання, які передбачені програмою навчання виконані в повному обсязі, відмінна робота без помилок або з однією незначною помилкою.
70 – 89	Добре	Теоретичний зміст курсу освоєний цілком, практичні навички роботи з освоєним матеріалом в основному сформовані, всі навчальні завдання, які передбачені програмою навчання виконані, якість виконання жодного з них не оцінено мінімальним числом балів, деякі види завдань виконані з помилками,робота з декількома незначними помилками, або з однією – двома значними помилками.
50 –69	Задовільно	Теоретичний зміст курсу освоєний не повністю, але прогалини не носять істотного характеру, необхідні практичні навички роботи з освоєним матеріалом в основному сформовані, більшість передбачених програмою навчання навчальних завдань виконано, деякі з виконаних завдань, містять помилки, робота з трьома значними помилками.
1–49	Незадовільно	Теоретичний зміст курсу не освоєно, необхідні практичні навички роботи не сформовані, всі виконані навчальні завдання містять грубі помилки, додаткова самостійна робота над матеріалом курсу не приведе до значимого підвищення якості виконання навчальних завдань, робота, що потребує повної переробки

10. Рекомендована література

Основна література

1. Ким Дж.О., Мьюллер Ч.У., Клекка У.Р., Енюков И.С. Факторный, дискриминантный и кластерный анализ. М.:Мир. - 1989. - 337с.
2. Халафян А.А. Учебник STATISTICA 6 Статистический анализ данных. М.:ООО Бином-Пресс. - 2007. 512с.
3. Доугерти К. Введение в эконометрику. Пер с англ. - М.:ИНФРА-М. 1999. 402с.
4. Павлов А.Н., Ермолаев Ю.М. Биоинформатика. Учебник для ВУЗов. М.: Гринлайт+, 2010. — 256 с.

Допоміжна література

1. Буреева Н.Н. Многомерный статистический анализ с использованием STATISTICA. Учебно-методические материалы. Нижний Новгород: ННГУ, 2007. – 112с.
2. Lowman M. A Practical Guide to Analytics for Governments: Using Big Data for Good. John Wiley & Sons. 2017. 224p.
3. 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. Введение в анализ больших данных http://www.vvsu.ru/files/B7DCFB93-1A78-4480-8EB8-3EFD0FD49033.pdf
2. Радченко И.А, Николаев И.Н. Технологии и инфраструктура Big Data. – СПб:Университет ИТМО, 2018. – 52 с. https://books.ifmo.ru/file/pdf/2326.pdf
3. Omar H., Jumaa 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
4. 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
5. Багато підручників з математичної статистики http://statosphere.ru/books-arch/multistat/69-k-muller.html
6. Файлы данных для обучения statlab.kubsu.ru