

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра прикладної математики

“ЗАТВЕРДЖУЮ”

Проректор

в науково-педагогічній роботі

Олександр ГОЛОВКО



“серпень” 2022 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Аналіз даних

(назва навчальної дисципліни)

рівень вищої освіти перший (бакалаврський)

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 обов'язкова
(обов'язкова / за вибором)

факультет математики і інформатики

2022 / 2023 навчальний рік

Програму рекомендовано до затвердження Вченою радою факультету математики і інформатики

“29” серпня 2022 року, протокол №7

РОЗРОБНИКИ ПРОГРАМИ: *Степанова Катерина Вадимівна*, кандидат фізико-математичних наук, доцент кафедри прикладної математики

Програму схвалено на засіданні кафедри Прикладної математики

Протокол від “29” серпня 2022 року №11

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 2022 року, протокол №1

Голова науково-методичної комісії факультету математики і інформатики



(підпис)

Ольга АНОЩЕНКО

(прізвище та ініціали)

ВСТУП

Програма навчальної дисципліни “Аналіз даних” складена відповідно до освітньо-професійної програми підготовки

бакалавр

(назва рівня вищої освіти)

Спеціальності (напряму) 113 Прикладна математика

спеціалізації _____

1. Опис навчальної дисципліни

1.1. Метою викладання навчальної дисципліни є надання майбутнім спеціалістам теоретичних знань та практичних навичок з методикою аналізу даних і основними принципами для розв’язання різних прикладних задач, що виникають при програмуванні, розробці та використанні сучасних інформаційних технологій.

1.2. Основні завдання вивчення дисципліни

- ознайомити студентів із основними поняттями аналізу даних;
- ознайомити з основними типами структур даних та їх використанням
- викласти основи аналізу даних в задачах моделювання й аналізу даних;
- вивчити використання методів аналізу даних при розв’язанні прикладних задач.

1.3. Кількість кредитів - 4

1.4. Загальна кількість годин - 120

1.5. Характеристика навчальної дисципліни	
Обов’язкова	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
4-й	-й
Семестр	
7-й	-й
Лекції	
32 год.	Год.
Практичні, семінарські заняття	
32 год.	Год.
Лабораторні заняття	
0 год.	Год.
Самостійна робота, у тому числі	
56 год.	Год.
у тому числі індивідуальні завдання	
6 год.	

1.6. Заплановані результати навчання. Студент повинен:

- Знати:* основні принципи та методи аналізу даних
основи синтаксису і особливості роботи в DataLore
- Вміти:* проводити аналіз даних;
застосовувати вивчені методи аналізу даних;
провести аналіз складності алгоритму, довести його коректність;
застосовувати сучасні програми, які містять алгоритми для аналізу даних.
- Володіти:* методами аналізу даних.

2. Тематичний план навчальної дисципліни

Тема 1. Вступ до аналізу даних.

Загальна інформація. Різні типи даних. Ресурси пошуку даних. Сучасні методи аналізу даних. Особливості синтаксису при роботі з даними, підключення модулів та бібліотек в DataLore. Приклади програм.

Тема 2. Візуальний аналіз даних.

Інтернет технологій. Виконання візуального аналізу даних - Visual Mining. Характеристики засобів візуалізації даних. Методи візуалізації.

Тема 3. Інтелектуальний аналіз даних.

Добування даних – Data Mining та його практичне застосування. Основні методи Data Mining. Проблеми аналізу інформації із Web. Передбачувані моделі. Пошук web-документів за допомогою гіперпосилань. Дослідження web-ресурсів.

Тема 4. Аналіз текстової інформації та даних в реальному часі.

Задача аналізу текстів. Класифікація текстових документів. Задачі анотації текстів. Засоби аналізу текстової інформації. Пошук інформації в реальному часі. Рекомендаційні системи. Інструменти Data Mining в реальному часі.

Тема 5. Елементи дисперсійного аналізу.

Однофакторний дисперсійний аналіз. Двофакторний дисперсійний аналіз. Особливості і приклади використання.

3. Структура навчальної дисципліни

Назви розділів і тем	Кількість годин					
	денна форма					
	усього	у тому числі				
л		п	лаб.	інд.	с. р.	
1	2	3	4	5	6	7
Тема 1. Вступ до аналізу даних	14	4	4			6
Тема 2. Візуальний аналіз даних	23	6	4			13

<i>Розрахунково-графічна робота</i>	2		2			
Тема 3. Інтелектуальний аналіз даних	40	12	10			18
<i>Розрахунково-графічна робота</i>	2		2			
Тема 4. Аналіз текстової інформації та даних в реальному часі	23	6	4			13
<i>Розрахунково-графічна робота</i>	2		2			
Тема 5. Елементи дисперсійного аналізу	14	4	4			6
Усього годин	120	32	32			56

4. Теми практичних занять

№ з/п	Назва теми	Кількість годин
1	Ресурси пошуку даних. Сучасні методи аналізу даних. Особливості синтаксису при роботі з даними, підключення модулів та бібліотек в DataLore. Приклади програм.	4
2	<i>Розрахунково-графічна робота</i>	2
3	Характеристики засобів візуалізації даних. Методи візуалізації VisualMining.	6
4	<i>Розрахунково-графічна робота</i>	2
5	Добування даних – Data Mining та його практичне застосування. Основні методи Data Mining. Проблеми аналізу інформації із Web. Передбачувані моделі. Пошук web-документів за допомогою гіперпосилань. Дослідження web-ресурсів.	6
6	Задача аналізу текстів. Класифікація текстових документів. Задачі анотації текстів. Засоби аналізу текстової інформації. Пошук інформації в реальному часі. Рекомендаційні системи. Інструменти Data Mining в реальному часі.	6
7	<i>Розрахунково-графічна робота</i>	2
8	Елементи дисперсійного аналізу	4
	Разом	32

5. Завдання для самостійної роботи

№ з/п	Види, зміст самостійної роботи	Кількість годин
-------	--------------------------------	-----------------

1	Виконання домашніх завдань з ознайомленням особливостей синтаксису, дослідження web- ресурсів.	6
2	Виконання домашніх завдань і написання найпростіших програм, проведення візуального аналізу даних досліджуваної моделі	13
3	Виконання домашніх завдань з опануванням практичного аспекту застосування Data Mining та його основних методів. Усвідомлення проблеми аналізу інформації із Web. Побудова та розгляд прикладів передбачуваних моделей.	18
4	Виконання домашніх завдань по аналізу текстової інформації та даних в реальному часі	13
5	Виконання домашніх завдань з дисперсійного аналізу	6
	Разом	56

6. Індивідуальні завдання

Не передбачені

7. Методи навчання

Частково-пошуковий, дослідницький, пояснювально-ілюстративний, репродуктивний. Проблемні лекції, пояснення, практичні роботи.

8. Методи контролю

- перевірка виконання домашніх завдань
- поточне опитування за лекційним матеріалом
- перевірка індивідуальних завдань
- перевірка звітів-результатів самостійної роботи студентів
- перевірка розрахунково-графічних робіт
- перевірка залікової роботи

9. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання						Залікова робота	Сума
T1	T2	T3	T4	T5	Графічно-розрахункові роботи, передбачені навчальним планом	Разом	
6	6	6	6	6	3 x 10	60	40
							100

T1, ..., T5 – теми.

Критерії оцінювання залікових робіт з курсу «Аналіз даних»

Білет складається з 2 теоретичних питань.

Максимальна оцінка за кожне питання 20 балів, разом за обидва - 40 балів.

Для отримання максимальної оцінки (20 балів) за одне питання потрібно сформулювати відповідні означення, основні поняттями аналізу даних, провести аналіз даних та/або алгоритми і обов'язково навести приклади. Замість прикладів можуть бути наведені міркування щодо змісту відповідних понять і результатів, проведена візуалізація (тобто надані інтерактивні графіки в DataStudio або plot line or pie chart; plot bar chart; plot box-plot в DataLore) в результаті проведеного аналізу даних. Викладення основних концепцій застосування сучасних програм, які містять алгоритми для аналізу даних та знання синтаксису і особливостей роботи в DataLore є обов'язковими.

Формулювання неповне або зі значними помилками: 5-15 балів.

Формулювання без прикладів або міркувань: 10-15 балів.

Шкала оцінювання (дворівнева)

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка
50-100	зараховано
1-49	не зараховано

10. Рекомендована література

Основна література

1. Барковський В.В. Теорія ймовірності та математична статистика. Навчальний посібник. – К: Центр учбової літератури, 2017. – 424 с.
2. Ганна Кармелюк. Теорія ймовірності та математична статистика. Посібник з розв'язування задач: Навчальний посібник рекомендовано МОН України. – К: Центр учбової літератури, 2017. – 576 с.
3. Галина Квасниця, Микола Притула, Ольга Прядко. Теорія ймовірностей та математична статистика, Частина 1 (Теорія ймовірностей). – Видавництво ЛНУ ім. Івана Франка, 2019. – 150 с.
4. Жлуктенко В. І., Наконечний С. І., Савіна С. С. Теорія ймовірностей і математична статистика: Навч.-метод. посібник: У 2-х ч. – Ч. II. Математична статистика. – К.: КНЕУ, 2001. – 336 с.
5. К. В. Степанова, Е. Ю. Железнякова, І. Л. Лебедєва, Л. О. Норік. Лабораторний практикум з навчальної дисципліни «Теорія ймовірностей та математична статистика» : навчальний посібник. – Харків: ХНЕУ ім. С.Кузнеця, 2016.– 184 с. – Режим доступу: <http://www.repository.hneu.edu.ua/jspui/handle/123456789/14255>

6. Литвин В.В., Нікольський Ю.В., Пасічник В.В. Аналіз даних та знань. Навчальний посібник. – Магнолія 2006, Серія: Комп'ютинг, 2021. – 276 с.
7. Черняк О.І. Інтелектуальний аналіз даних: підручник. – К: Знання, 2014. –599с.
8. Данильченко О.М., Данильченко А.О. Інтелектуальний аналіз даних: Навч. посібник. – Житомир: ЖДТУ, 2009. – 405 с.
9. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. – Запоріжжя: КПУ, 2011. – 268 с.
10. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посібник. — К.: КНЕУ, 2007. — 376 с.

Допоміжна література

1. A. Bryman, Handbook of Data Analysis. Pub. by Sage Publications, 2009, 728 p.
2. Daniel T. Larose, Data Mining methods and models. Wiley, 2006, p.322
3. Daniel T. Larose, Discovering knowledge in data. Wiley, 2005, p.222
4. Alexander Paprotny, Michael Thess, Realtime data mining. Springer, 2013, p.313

11. Посилання на інформаційні ресурси в Інтернеті

1. Recommendation system video lectures Xavier Amatrian
<http://technocalifornia.blogspot.com/2014/08/introduction-to-recommender-systems-4.html>
2. Data Science
<https://opendatascience.com>
3. Knowledge discovering
<http://www.kdnuggets.com>
4. Neo4j DataBase
<https://neo4j.com>
5. RapidMiner
<https://rapidminer.com>