

**НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ ПРИКЛАДНОЇ МАТЕМАТИКИ І МЕХАНІКИ**

«ЗАТВЕРДЖУЮ»

Директор ІПММ НАН України
чл.-кор. НАН України



І.І. Скрипнік

«23» серпня 2021 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Сучасні прикладні методи інтелектуального аналізу даних та знань»

підготовки здобувачів ступеня вищої освіти
треть (освітньо-науковий)
доктора філософії

галузь знань 11 Математика та статистика

спеціальність 113 Прикладна математика

Робоча програма навчальної дисципліни «Сучасні прикладні методи інтелектуального аналізу даних та знань» для здобувачів наукового ступеня доктора філософії за спеціальністю 113 «Прикладна математика»

Розробник:

Сенченко Олексій Сергійович, старший науковий співробітник відділу теорії керуючих систем, кандидат фізико-математичних наук, доцент

Робочу програму схвалено на засіданні відділу теорії керуючих систем

Протокол № 2 від 26. 08. 2021 року.

Завідувач відділу теорії керуючих систем  Кононов Ю.М.

Схвалено Науково-методичною радою Інституту прикладної математики і механіки НАН України

Протокол № 4 від 30 серпня 2021 року.

Голова науково-методичної ради



Несмелова О.В.

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни
Кількість кредитів - 4	Галузь знань: 11 «Математика та статистика»	дисципліна вільного вибору аспіранта
Загальна кількість годин – 120	Спеціальність: 113 «Прикладна математика»	Рік підготовки:
		1-2
		Семестр
		2, 3
		Лекції
Тижневих годин: аудиторних – 1,5 самостійної роботи – 1,7	Освітньо-кваліфікаційний рівень: доктор філософії	28 год
		Практичні
		28 год
		Самостійна робота
		64 год
		Вид контролю: іспит

2. Мета, завдання та очікувані результати навчальної дисципліни.

Метою навчальної дисципліни «Сучасні прикладні методи інтелектуального аналізу даних та знань» є послідовне викладення методів сучасної обробки даних – аналітичного дослідження великих масивів даних з метою виявлення у них нових, раніше невідомих, нетривіальних, практично корисних знань і закономірностей, необхідних для прийняття рішень у різноманітних сферах людської діяльності, розгляд практичних прикладів застосування інтелектуального аналізу даних та огляд методів, програмних продуктів і різних інструментальних засобів, які використовуються в аналізі даних.

Завдання навчальної дисципліни:

- ознайомлення та оволодіння класичними та сучасними методами інтелектуального аналізу даних;
- оволодіння теоретичними положеннями та основними застосуваннями лінійної алгебри, теорії баз даних, теорії ймовірності, математичної статистики, дослідження операцій у цих методах;
- отримання практичних навичок з використання інструментальних засобів інтелектуального аналізу даних при розв'язанні прикладних задач та вірної інтерпретації отриманих результатів.

Очікувані результати навчальної дисципліни

Загальні компетентності:

- ЗК1. Аналіз і синтез. Здатність до аналізу та синтезу на основі логічних аргументів та перевірених фактів; навички управління інформацією (уміння знаходити та аналізувати інформацію з різних джерел).
- ЗК3. Гнучкість мислення. Набуття гнучкого способу мислення, який дає можливість зрозуміти й розв'язати проблеми та задачі, зберігаючи при цьому критичне відношення до усталених наукових концепцій.
- ЗК4. Креативність. Потенціал креативності у генеруванні ідей та досягненні наукових цілей.
- ЗК5. Групова робота. Здатність до взаємодії (робота в команді); здатність працювати в міждисциплінарній команді.
- ЗК7. Міжнародний кругозір. Здатність працювати в міжнародному середовищі, ставитися із повагою до національних та культурних традицій, способів роботи інших членів групи.
- ЗК8. Популяризаційні навички. Вміння спілкуватися із нефаківцями; здатність провести усну презентацію та написати зрозумілу статтю за результатами проведених досліджень, а також щодо сучасних концепцій у прикладній математиці для нефаківців у цій галузі; здатність спілкуватися з експертами з інших галузей.
- ЗК11. Етичні установки. Етика поведінки в наукових дослідженнях; турбота про якість результату.

- ЗК12. Особисті якості. Здатність до навчання, саморозвитку та самовдосконалення протягом життя; визначеність і наполегливість щодо поставлених завдань і взятих обов'язків.

Фахові компетентності:

- ФК2. Здатність застосовувати сучасну методологію, загальні та спеціальні методи наукового дослідження за спеціальністю та обраною спеціалізацією.
- ФК 5. Уміння будувати адекватні математичні моделі, досліджувати побудовані моделі та визначати можливості їх застосування.
- ФК 8. Здатність здійснювати дослідження прикладних математичних задач з використанням спеціальних математичних пакетів та аналізувати отримані дані.
- ФК18. Здатність ефективно аналізувати великі набори різномірних даних із використанням методів машинного навчання, статистики, штучного інтелекту тощо.

Програмні результати навчання:

- РН 1. Знати та розуміти традиційні та передові тенденції в галузі прикладної математики, етапи історичного розвитку математичних знань і парадигм, формування поглибленої системи наукових знань за напрямом досліджень.
- РН 3. Знати основні принципи, форми та методи проведення наукових досліджень в галузі прикладної математики, володіти математичним апаратом в об'ємі, достатньому для професійної діяльності.
- РН 4. Вміти будувати та досліджувати математичні моделі фізичних, соціально-економічних, технічних, біологічних тощо процесів, систем, процесів, формалізувати проблеми, описані природною мовою, за допомогою математичних методів, застосовувати загальні підходи до математичного моделювання конкретних процесів.
- РН5. Мати навички використання загальних офісних та спеціалізованих програмних засобів за напрямом професійної діяльності, вміти користуватися спеціалізованими інтернет-ресурсами.
- РН9. Вміти організувати неперервний власний саморозвиток і самовдосконалення, набуття гнучкого способу мислення при творчому дослідженні, в тому числі прикладних математичних проблем.
- РН 15. Уміти здійснювати збір, опрацювання, аналіз, систематизацію, захист спеціалізованої науково-технічної інформації з різних джерел, уникаючи при цьому академічної недоброчесності, робити візуалізацію аналізу інформації.
- РН 17. Знати основні чисельні методи та алгоритми розв'язання задач прикладної та обчислювальної математики.

- РН 18. Знати та вміти досліджувати основні математичні моделі, що використовуються для описання та дослідження дискретних систем різних типів, реалізовувати алгоритми моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити експерименти.

За результатами вивчення навчальної дисципліни здобувач доктора філософії повинен:

знати:

- базові визначення, змістову і математичну постановку основних задач аналізу даних та способи їх розв'язку;
- сутність і властивості основних алгоритмів інтелектуального аналізу даних;
- основи організації та технології комп'ютерної обробки експериментальних і статистичних даних.

вміти:

- обґрунтовувати вибір засобів і методів для розв'язку конкретних задач аналізу даних;
- застосовувати вивчені методи аналізу даних;
- представляти результати аналізу в зручному для сприйняття вигляді.

3. Програма навчальної дисципліни

Тема 1. Системи підтримки рішень (СПР). [1:15-28; 2:13-26]

1. Структура СПР. Класифікація задач аналізу даних.
2. Бази даних. 12 правил Кодда для реляційних баз даних.
3. OLTP-системи. Основні суперечності між OLTP- та СПР-системами.

Тема 2. Сховища даних. [1:29-49; 2:27-47; 3:95-140; Д2:83-120]

1. Означення та властивості сховища даних.
2. Організація сховища даних. Вітрина даних.
3. Основні операції з даними при переміщенні їх до сховища.

Тема 3. OLAP-системи. [1: 50-67; 2:49-66; 3:144-192; Д2:135-149, 156-164]

1. Операції з даними, представленими у багатовимірній моделі. Означення OLAP-систем.
2. Правила Кодда для OLAP-систем.
3. Архітектура OLAP-систем.

Тема 4. Інтелектуальний аналіз даних. [1: 68-84, 90-101; 2:67-83, 99-104; 3:18-52; 5: 15-19, Д1:61-82, Д2: 83-120, Д5: 14-32, 246-306]

1. Задачі Data Mining.
2. Практичне застосування Data Mining.
3. Моделі Data Mining.
4. Засоби Data Mining.
5. Процес виявлення знань.

Тема 5. «М'які» обчислення у Data Mining. [1: 84-90; 2:83-89; 3: 199-293, 431-522, 545-634; 4: 109-135, 210-259, 5: 96-99, 6: 38-41, 7: 42-95, Д3: 284-311, Д5: 131-183]

1. Еволюційні технології та генетичні алгоритми.
2. Нечітка логіка.
3. Нейронні мережі.

Тема 6. Задачі класифікації даних у Data Mining. [1:102-136; 2:95-128; 3: 640-690; 5: 39-44, 46-60; 7:111-130; Д2: 327-438; Д3: 312-328; Д4: 62-104]

1. Способи представлення функціональних залежностей у задачах класифікації.
2. 1R-алгоритм та метод Naive Bayes.
3. Метод опорних векторів.
4. Методика «розділяй та володарюй».
5. Алгоритми ID3 та C4.5.
6. Алгоритм покриття.
7. Лінійні та нелінійні алгоритми класифікації.

Тема 7. Задачі регресії у Data Mining. [4:107-109; 5: 44-46; Д1: 243-253; Д2: 99-105; Д4: 105-108, 161-165; Д5: 65-78]

1. Відмінності між задачами класифікації та регресії.
2. Побудова дерева регресії.

Тема 8. Аналіз часових рядів. [1:136-140; 5: 20-38; 6: 42-48; 7:136-142; Д1: 311-312; Д2: 586-591]

1. Поняття часового ряду. Етапи прогнозування часових рядів.
2. Класифікація та екстраполяція часових рядів.
3. Лінійний дискримінантний аналіз часових рядів (LDA).
4. Квадратичний дискримінантний аналіз часових рядів (QDA).

Тема 9. Алгоритми кластеризації у Data Mining. [1:159-191; 2: 149-208; 3: 691-750; 5:61-95; 6: 8-25; 7: 97-111; Д1: 136-139, 254-270; Д3: 50-76; Д4: 256-272]

1. Кластеризація у Data Mining.
2. Ієрархічні алгоритми кластеризації.
3. Неієрархічні алгоритми кластеризації.
4. Кластеризація із використанням нечітких відношень.

Тема 10. Пошук асоціативних правил у Data Mining. [1:140-158; 2: 129-148; 3: 322-430; 4:75-87; Д1: 112-118]

1. Постановка задачі пошуку асоціативних правил.
2. Алгоритм Apriori та його різновиди.
3. Алгоритм Eclat.
4. Алгоритми на основі множини вузлів.

Тема 11. Елементи комп'ютерної лінгвістики. Text Mining. [1:211-241; 5: 244-276; 6: 42-48; 7:136-142; Д1: 351-356]

1. Задачі Text Mining.
2. Методи класифікації та кластеризації текстових документів.
3. Засоби аналізу текстової інформації.
4. Емпіричні закономірності Ципфа, Бредфорда, Хіпса.
5. Онтології.

Тема 12. Data Mining у реальному часі. [1:325-349]

1. Концепція Data Mining у реальному часі.
2. Рекомендаційні системи.
3. Інструменти Data Mining у реальному часі.

4. Структура навчальної дисципліни

Назва теми	Кількість годин			
	Усього	у тому числі		
		Лекції	Практичні заняття	Самостійна робота
1	2	3	4	5
Тема 1. Системи підтримки рішень	7	2	2	3
Тема 2. Сховища даних	7	2	2	3
Тема 3. OLAP-системи	8	2	2	6
Тема 4. Інтелектуальний аналіз даних	9	2	2	5
Тема 5. «М'які» обчислення у Data Mining	12	2	2	8
Тема 6. Задачі класифікації даних у Data Mining	14	4	4	6
Тема 7. Задачі регресії у Data Mining	10	2	2	6
Тема 8. Аналіз часових рядів	9	2	2	5
Тема 9. Алгоритми кластеризації у Data Mining.	10	2	2	6
Тема 10. Пошук асоціативних правил у Data Mining.	10	2	2	6
Тема 11. Елементи комп'ютерної лінгвістики. Text Mining	14	4	4	6
Тема 12. Data Mining у реальному часі	10	2	2	4
Усього годин	120	28	28	64

5. Теми практичних занять

№ з/п	Назва теми	Кількість годин
1	Системи підтримки рішень.	2
2	Сховища даних.	2
3	OLAP-системи.	2
4	Інтелектуальний аналіз даних.	2
5	«М'які» обчислення у Data Mining.	2
6	Задачі класифікації даних у Data Mining.	4
7	Задачі регресії у Data Mining.	2
8	Аналіз часових рядів.	2
9	Алгоритми кластеризації у Data Mining.	2
10	Пошук асоціативних правил у Data Mining.	2
11	Елементи комп'ютерної лінгвістики. Text Mining.	4
12	Data Mining у реальному часі.	2
	Разом	28

6. Теми для самостійної роботи

№ з/п	Назва теми	Кількість годин
1	Системи підтримки рішень	3
2	Сховища даних	3
3	OLAP-системи	6
4	Інтелектуальний аналіз даних	5
5	«М'які» обчислення у Data Mining	8
6	Задачі класифікації даних у Data Mining	6
7	Задачі регресії у Data Mining	6
8	Аналіз часових рядів	5
9	Алгоритми кластеризації у Data Mining	6
10	Пошук асоціативних правил у Data Mining	6
11	Елементи комп'ютерної лінгвістики. Text Mining	6
12	Data Mining у реальному часі	4
	Разом	64

7. Методи навчання

При вивченні дисципліни використовуються:

1. Дидактичні методи – лекції з використанням мультимедійних презентацій.
2. Практичні методи: практичні роботи з використанням прикладного програмного забезпечення
3. Метод самостійного навчання.
4. Активні методи: експрес опитування.
5. Словесні методи навчання: лекції, бесіди

8. Методи контролю

Використовуються наступні методи контролю:

1. Поточний контроль знань:
 - усне опитування;
 - виконання практичних робіт;
 - виконання дослідницьких завдань.
2. Підсумковий контроль знань (іспит):
 - екзаменаційні білети.

9. Критерії оцінювання результатів навчання

Розподіл балів, які отримує аспірант

Аудиторна робота												Самостійна робота	Підсумковий контроль (іспит)	Загальна кількість отриманих балів
теми														
T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12			
4	4	4	4	4	5	4	4	4	4	5	4	10	40	100

Шкала оцінювання

Сума балів за всі види навчальної діяльності	Оцінка за національною шкалою
	для іспиту
90-100	«відмінно» - здобувач повно і всебічно розкриває питання теми, вільно оперує поняттями і термінологією, демонструє глибокі знання джерел, має власну точку зору стосовно відповідної теми і може аргументовано її доводити
75-89	«добре» - загалом рівень знань здобувачів відповідає викладеному вище, але мають місце деякі упущення при виконанні завдань, обґрунтування неточні, не підтверджуються достатньо обґрунтованими доказами
60-74	«задовільно» - здобувач розкрив питання, в загальних рисах, розуміє їх суть, намагається робити висновки, але при цьому припускається грубих помилок, матеріал викладає нелогічно і не самостійно
35-59	«незадовільно» - з можливістю повторного складання, здобувач не в змозі дати відповідь на поставлене запитання або відповідь неправильна, не розуміє суті питання, не може зробити висновки
0-34	«незадовільно» - з обов'язковим повторним вивченням дисципліни

10. Орієнтовний перелік питань до підсумкового контролю (іспиту)

1. Структура систем підтримки рішень. Класифікація задач аналізу даних.
2. Бази даних. 12 правил Кодда для реляційних баз даних.
3. OLTP-системи. Основні суперечності між OLTP- та СПР-системами.
4. Означення та властивості сховища даних.
5. Організація сховища даних. Вітрина даних.
6. Основні операції з даними при переміщенні їх до сховища.
7. Операції з даними, представленими у багатовимірній моделі. Означення OLAP-систем.
8. Правила Кодда для OLAP-систем.
9. Архітектура OLAP-систем.
10. Задачі Data Mining.
11. Практичне застосування Data Mining.
12. Моделі Data Mining.
13. Засоби Data Mining.
14. Процес виявлення знань.
15. Поняття та можливості нейрокомп'ютерних технологій.
16. Архітектура нейронних мереж.
17. Нейронні мережі Хопфілда та Кохонена.
18. Програмні засоби реалізації нейрокомп'ютерних технологій.

19. Концептуальні засади еволюційної теорії.
20. Основні положення теорії генетичних алгоритмів.
21. Моделі генетичних алгоритмів.
22. Програмне забезпечення та сфері застосування генетичних алгоритмів.
23. Концепція нечітких обчислень.
24. Програмне забезпечення нечітких методів.
25. Сучасна практика застосування нечітких методів.
26. Способи представлення функціональних залежностей у задачах класифікації.
27. 1R-алгоритм та метод Naive Bayes.
28. Метод опорних векторів.
29. Методика «розділяй та володарюй».
30. Алгоритми ID3 та C4.5.
31. Алгоритм покриття.
32. Лінійні та нелінійні алгоритми класифікації.
33. Відмінності між задачами класифікації та регресії.
34. Побудова дерева регресії.
35. Поняття часового ряду. Етапи прогнозування часових рядів.
36. Класифікація та екстраполяція часових рядів.
37. Лінійний дискримінантний аналіз часових рядів (LDA).
38. Квадратичний дискримінантний аналіз часових рядів (QDA).
39. Кластеризація у Data Mining.
40. Ієрархічні алгоритми кластеризації.
41. Неієрархічні алгоритми кластеризації.
42. Кластеризація із використанням нечітких відношень.
43. Постановка задачі пошуку асоціативних правил.
44. Алгоритм пошуку асоціативних правил Apriori та його різновиди.
45. Алгоритм пошуку асоціативних правил Eclat.
46. Алгоритми пошуку асоціативних правил на основі множини вузлів.
47. Задачі Text Mining.
48. Методи класифікації та кластеризації текстових документів.
49. Засоби аналізу текстової інформації.
50. Емпіричні закономірності Ципфа, Бредфорда, Хіпса.
51. Онтології.
52. Концепція Data Mining у реальному часі.
53. Рекомендаційні системи.
54. Інструменти Data Mining у реальному часі.

11. Рекомендована література

Основна література дисципліни

1. Анализ данных и процессов : учеб. пособие : 3-е изд., перераб. и доп. / А.А. Барсегян и др. Санкт-Петербург : БХВ-Петербург, 2009. 512 с.
2. Методы и модели анализа данных: OLAP и Data Mining. / А.А. Барсегян и др. Санкт-Петербург : БХВ-Петербург, 2004. 336 с.
3. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних : підручник. Київ : Знання, 2010. 842 с.
4. Олійник А.О., Субботін С.О., Олійник О.О. Інтелектуальний аналіз даних : навчальний посібник. Запоріжжя : ЗНТУ, 2012. 278 с.
5. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки : навчальний посібник. Київ : ІСЗЗІ КПІ ім. Ігоря Сікорського, 2018. 297 с.
6. Применение методов интеллектуального анализа данных и процессов: практикум / сост. Т. В. Афанасьева. Ульяновск : УлГТУ, 2018. 51 с.
7. Шумейко А.А., Сотник С.Л. Интеллектуальный анализ данных (Введение в Data Mining) : учеб. пособ. Днепропетровск, 2012. 212 с.

Допоміжна література дисципліни

1. Witten I.H. & Frank E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd edition. San Francisco: Morgan Kaufmann Publishers.
2. Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques, third edition* (3rd ed.). Waltham, Mass.: Morgan Kaufmann Publishers.
3. Сегаран. Т. Программируем коллективный разум : пер. с англ. Санкт-Петербург : Символ-Плюс, 2008. 368 с.
4. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : пер. с англ. Москва : ДМК Пресс, 2015. 400 с.
5. Дюк В., Самойленко А. Data Mining : учебный курс. Санкт-Петербург : Питер, 2001. 368 с.