

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ

Кафедра теоретичної кібернетики

«ЗАТВЕРДЖУЮ»
Заступник декана
з навчальної роботи

Олена КАШПУР
« 12 » _____ 2021 року

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
ІНТЕЛЕКТУАЛЬНА ОБРОБКА ДАНИХ**

для студентів

галузь знань **12 «Інформаційні технології»**
(шифр і назва)

спеціальність **122 «Комп'ютерні науки»**
(шифр і назва спеціальності)

освітній рівень **бакалавр**
(молодший бакалавр, бакалавр, магістр)

освітня програма **«Інформатика»**
(назва освітньої програми)

вибірковий блок **«Інформаційні технології та системи»**

вид дисципліни **вибіркова**

Форма навчання	денна
Навчальний рік	2022/2023
Семестр	7
Кількість кредитів ECTS	4
Мова викладання, навчання та оцінювання	українська
Форма заключного контролю	іспит

Викладачі: **Анатолій ПАШКО, д.ф.-м.н., проф. (лекції)**

Пролонговано: на 20__/20__ н.р. _____ (_____) «__» __20__р.
(підпис, ПІБ, дата)


на 20__/20__ н.р. _____ (_____) «__» __20__р.
(підпис, ПІБ, дата)

КИЇВ – 2021

Розробники: Анатолій ПАШКО, д.ф.-м.н., професор, професор кафедри теоретичної кібернетики

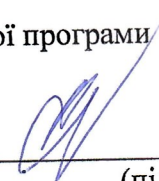
ЗАТВЕРДЖЕНО

Зав. кафедри теоретичної кібернетики


Юрій КРАК
(підпис)

Протокол № 7 від « 8 » листопада 2021 р.

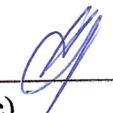
Схвалено Гарантом освітньо-професійної програми «Інформатика»


Людмила ОМЕЛЬЧУК
(підпис)

« 11 » листопада 2021 р.

Схвалено науково-методичною комісією факультету комп'ютерних наук та кібернетики

Протокол від « 11 » листопада 2021 року № 7

Голова науково-методичної комісії 
(підпис) Людмила ОМЕЛЬЧУК

« 11 » листопада 2021 року

1 Мета дисципліни – вивчення основних та найбільш перспективних напрямків аналізу даних: зберігання інформації, оперативний і інтелектуальний аналіз, а також методів та алгоритмів інтелектуального аналізу; знайомство з актуальними питаннями, що постають при розробці програмних продуктів, що обробляють великі обсяги даних.

2 Попередні вимоги до опанування або вибору навчальної дисципліни

Знати: дискретну математику, математичну логіку, теорію алгоритмів, структури даних і алгоритми, парадигми програмування та основи програмування в об'ємі стандартних університетських курсів.

Вміти: застосовувати знання з вказаних вище дисциплін до розв'язання задач, вміти писати програми мовою високого рівня.

Володіти: елементарними навичками використання логіко-математичної символіки та основами програмування

3 Анотація навчальної дисципліни

Викладається у 7 семестрі 4 курсу в обсязі – 120 год. (4 кредити ECTS), зокрема: лекції – 42 год., консультації – 2 год., самостійна робота – 76 год.

Протягом вивчення студенти мають опанувати основні методи та моделі інтелектуального аналізу даних та засоби їх реалізації, зокрема оволодіти алгоритмами обробки потокових даних, навчитися аналізувати та уникати сучасних проблем, пов'язаних із збиранням та обробленням інформації.

Отримані знання дозволять ефективно будувати програмні продукти, що комплексно використовують аналітичні, інформаційні та комунікаційні технології обробки та аналізу даних.

4 Завдання (навчальні цілі)

Набуття поглиблених фахових компетентностей. Зокрема, розвивати:

СК2. Здатність до виявлення закономірностей випадкових явищ, застосування методів статистичної обробки даних та оцінювання стохастичних процесів реального світу.

СК11. Здатність до інтелектуального багатовимірного аналізу даних та їхньої оперативної аналітичної обробки з візуалізацією результатів аналізу в процесі розв'язування прикладних задач у галузі комп'ютерних наук.

СК12. Здатність забезпечити організацію обчислювальних процесів в інформаційних системах різного призначення з урахуванням архітектури, конфігурування, показників результативності функціонування операційних систем і системного програмного забезпечення.

СК18.3. Здатність використовувати технології штучного інтелекту та взаємодії «людина-комп'ютер».

5 Результати навчання за дисципліною

Результат навчання (РН) (1 – знати; 2 – вміти; 3 – комунікація; 4 – автономність та відповідальність)		Форми викладання та навчання	Методи оцінювання та пороговий критерій оцінювання (за необхідності)	Відсоток у підсумковій оцінці з дисципліни
Код	Результат навчання			
РН1.1	Знати основні методи та алгоритми інтелектуальної обробки даних, їх реалізацію сучасними мовами програмування	Лекція, самостійна робота	Контрольна робота (КР), ПЗ, іспит	20
РН1.2	Знати принципи застосування основних сучасних методів та алгоритмів при розробці елементів програмного забезпечення для інтелектуальної обробки даних	Лекція, самостійна робота	Контрольна робота (КР), ПЗ іспит	20
РН2.1	Вміти застосовувати сучасні алгоритми при розробці елементів програмного забезпечення	Лекція, самостійна робота	КР, ПЗ, іспит	20
РН3.1	Обґрунтовувати власний погляд на задачу, спілкуватися з колегами з питань	Самостійна робота	КР	10

	<i>проектування та розробки програм, складати письмові звіти</i>			
РН4.1	<i>Організувати свою самостійну роботу для досягнення результату</i>	<i>Самостійна робота</i>	<i>КР, ПЗ</i>	<i>10</i>
РН4.2	<i>Відповідально ставитися до виконуваних робіт, нести відповідальність за їх якість</i>	<i>Лекція, самостійна робота</i>	<i>КР, ПЗ, іспит</i>	<i>20</i>

6 Співвідношення результатів навчання дисципліни з програмними результатами навчання

Результати навчання дисципліни	РН1.1	РН1.2	РН2.1	РН3.1	РН4.1	РН4.2
	Програмні результати навчання					
<i>(з опису освітньої програми)</i>						
ПРН12. Вміти застосовувати методи та алгоритми інтелектуального аналізу даних для задач класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірної аналізу даних на основі використання технологій DataMining, TextMining, WebMining.			+	+		
ПРН19.3. Знати алгоритми аналізу інформації та вміти застосовувати їх у розв'язанні практичних задач.	+	+		+	+	+
ПРН21.3. Знати технології штучного інтелекту та вміти застосовувати їх у розв'язанні практичних задач.		+	+	+	+	+

7 Схема формування оцінки

7.1 Форми оцінювання студентів:

- семестрове оцінювання:

1. Контрольна робота 1 (за частиною 1): РН1.1, РН1.2, РН2.1, – 20 балів/ 12 балів.
2. Контрольна робота 2 (за частиною 2): РН 1.1., РН1.2, РН2.1, РН 3.1 – 20 балів/ 12 балів.
3. Практичне завдання за частиною 2 (програмна реалізація алгоритму з частини 2 із застосуванням до набору тестових даних, отриманням чисельного результату та складанням звіту): РН 1.1., РН1.2, РН2.1, РН 3.1, РН4.1, РН4.2 — 20 балів/ 12 балів.

Типова контрольна робота 1 складається з теоретичних та практичних завдань (з відкритими та закритими відповідями) за матеріалом розділу 1.

Матеріал, що виноситься на контрольну роботу 1:
див. запитання 1-24 для підготовки до оцінювання.

Типова контрольна робота 2 складається з теоретичних та практичних завдань (з відкритими та закритими відповідями) за матеріалом розділів 1 та 2.

Матеріал, що виноситься на контрольну роботу 2:
див. запитання 1-62 для підготовки до оцінювання.

У разі неякісного виконання практичного завдання викладач має право її не зарахувати або знизити бали.

- підсумкове оцінювання (у формі іспиту):

- максимальна кількість балів які можуть бути отримані студентом: 40;
- результати навчання, які оцінюються: РН1.1, РН1.2, РН2.1;
- форма проведення: письмова
- види завдань: одне теоретичне питання (40%, питання 1-62), дві задачі (60%).

Студент допускається до іспиту, якщо в семестрі набрав не менше ніж 20 балів.

Питання до оцінювання

1. Дайте визначення поняттю даних. Які типи даних ви знаєте?
2. Що собою являє об'єкт та атрибут в табличних даних?
3. Які типи шкал ви знаєте? В чому особливість кожної з них?
4. Що собою являє попередній аналіз даних?
5. Методи дисперсійного аналізу.
6. Особливості кореляційного аналізу.
7. Метод головних компонент.
8. Зменшення розмірності.
9. Множинна перевірка гіпотез
10. Ефект множинних порівнянь.
11. Міри помилки першого роду при множинній перевірці.
12. Які властивості методів Data Mining ви знаєте?
13. Що таке масштабованість системи?
14. Основні поняття інтелектуальної обробки даних (ІОД).
15. Сутність аналітичних технологій, поняття ІОД.
16. Етапи та методи пошуку нових знань.
17. Основні моделі інтелектуальних обчислень.
18. Засоби програмної підтримки ІОД.
19. Інтелектуальний аналіз даних – Data Mining.
20. Класифікація задач Data Mining.
21. Процес виявлення знань. Основні етапи аналізу. Підготовка початкових даних.
22. Основні поняття Rules Mining.
23. Програмні засоби пошуку асоціативних правил.
24. Практичний аспект застосування технології асоціативних правил.
25. Формальна постановка задачі кластеризації.
26. Представлення результатів. Базові алгоритми кластеризації.
27. Виконання візуального аналізу даних.
28. Характеристики засобів візуалізації даних.
29. Методи візуалізації. Методи геометричних перетворень.
30. Задача та етапи аналізу текстів Text Mining.
31. Загальний опис процесу отримання понять з тексту.
32. Методи класифікації текстових документів.
33. Методи кластеризації текстових документів: ієрархічні, бінарні.
34. Представлення текстових документів.
35. Задача анотування текстів.
36. Засоби аналізу текстової інформації.
37. Постановка завдання класифікації.
38. Визначення класифікатора.
39. Способи завдання класифікатора.
40. Якість класифікатора.
41. Баєсівський класифікатор.
42. Мінімаксний класифікатор.
43. Класифікатор Неймана-Пірсона.
44. Відношення правдоподібності.
45. Баєсівський класифікатор для нормально розподілених векторів ознак.
46. Баєсівський класифікатор для розпізнавання бінарних векторів ознак.
47. Обчислення ймовірностей помилкової класифікації.
48. Нормальний дискримінантний аналіз.
49. Лінійний дискримінант Фішера.
50. Оптимальна розділяюча гіперплощина.

51. Метод опорних векторів.
52. Машинне навчання і аналіз даних.
53. Дедуктивні логічні міркування і висновки.
54. Ймовірнісне виведення.
55. Баєсівська мережа.
56. Марковська мережа.
57. Приховані марковські моделі.
58. Дворівнева баєсівська ПММ.
59. Пуассонівська ПММ.
60. Алгоритми Монте-Карло за схемою марковського ланцюга.
61. Алгоритм Метрополіса-Гастингса.
62. Машинне навчання у випадку великої розмірності

7.2 Організація оцінювання

Терміни проведення форм оцінювання:

1. Контрольна робота 1: через тиждень після лекції №10.
2. Контрольна робота 2: після лекції №21.
3. Захист лабораторної роботи 1: до 14 тижня семестру включно.

Тема практичного завдання вибирається з переліку тем лекцій та узгоджується з викладачем не пізніше 6-го тижня семестру включно. Якщо під час захисту виявляється, що студент не до кінця розуміє використаний метод, код або погано в ньому орієнтується чи не розуміє використані синтаксичні елементи мови, зміст та призначення частин коду, то практичне завдання отримує нуль балів.

У випадку відсутності студента з поважних та документарно підтверджених причин відпрацювання та перездачі контрольних робіт здійснюються у відповідності до нормативних документів Університету, але не пізніше останнього робочого дня теоретичного навчання в семестрі.

Після завершення теоретичного навчання в семестрі і до офіційно призначеного перескладання жодні контрольні заходи не проводяться та жодні завдання не приймаються.

7.3 Шкала відповідності оцінок

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59

8. Структура навчальної дисципліни. Тематичний план лекцій

№ п/п	Назва теми	Кількість годин	
		Лекції	Самостійна робота
Частина 1. Технології інтелектуальної обробки даних.			
1	Тема 1. Інформація і знання. Дані, класифікація даних. <i>Самостійна робота:</i> Описова та варіаційна статистика.	2	2
2	Тема 2. Статистичний аналіз даних: перевірка гіпотез, дисперсійний аналіз, регресійний аналіз, кореляційний аналіз. <i>Самостійна робота:</i> Проведення статистичного аналізу в середовищі R Studio, Python.	2	6
3	Тема 3. Факторний аналіз. Метод головних компонент. <i>Самостійна робота:</i> Основні методи факторного аналізу: головних факторів, центроїдний, масимуму правдоподібності, контрастних груп та інші.	2	4
4	Тема 4. Множинна перевірка гіпотез. <i>Самостійна робота:</i> Методи вирішення проблеми множинних порівнянь.	2	4
5	Тема 5. Традиційні та сучасні технології інтелектуальної обробки даних. <i>Самостійна робота:</i> Сховища даних та OLAP-технології. Стандарти Data Mining.	2	4
6	Тема 6. Пошук асоціативних правил - Rules Mining та кластерний аналіз. <i>Самостійна робота:</i> Кластеризація по ГюстафсонуКесселю	4	4
7	Тема 7. Візуальний аналіз даних - Visual Mining. <i>Самостійна робота:</i> Характеристики сучасних засобів візуалізації даних	2	4
8	Тема 8. Аналіз текстової інформації – Text Mining. <i>Самостійна робота:</i> Використання машинного навчання до аналізу текстів.	4	6
	<i>Контрольна робота 1</i>		2
	Всього по частині 1	20	36
Частина 2. Інтелектуальний аналіз даних			
9	Тема 9. Задача класифікації. Оптимальні стратегії статистичної класифікації. <i>Самостійна робота:</i> наївний баєсівський класифікатор – визначення спаму	2	4
80	Тема 10. Баєсівська класифікація в типових задачах розпізнавання образів <i>Самостійна робота:</i> Реалізація баєсівської класифікації в мовах програмування.	2	4
91	Тема 11. Ймовірність помилкової класифікації <i>Самостійна робота:</i> Експериментальна оцінка ймовірності помилкової класифікації.	2	4
102	Тема 12. Дискримінантний аналіз <i>Самостійна робота:</i> Методи регуляризації	2	4

	коваріаційної матриці.		
113	Тема 13. Метод опорних векторів (SVM) <i>Самостійна робота:</i> Метод опорних векторів в задачах регресії.	2	4
124	Тема 14. Машинне навчання і аналіз даних. Два підходи <i>Самостійна робота:</i> Індуктивні та дедуктивні логічні міркування і висновки	2	4
135	Тема 15. Баєсівські та марковські мережі <i>Самостійна робота:</i> Марковські випадкові поля	2	4
146	Тема 16. Приховані марковські моделі <i>Самостійна робота:</i> Програмна реалізація прихованих марковських моделей.	4	4
157	Тема 17. Алгоритми Монте-Карло за схемою марковського ланцюга. <i>Самостійна робота:</i> Системи ймовірнісного програмування.	4	6
	<i>Контрольна робота 2</i>		2
	Всього по частині 2	22	40
	Всього	42	76

Загальний обсяг 120 год. (3 кредити ECTS), у тому числі:

Лекцій – 42 год.

Консультації – 2 год.

Самостійна робота – 76 год.

9. Рекомендовані джерела

Основні:

1. Марченко О.О. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики / О.О. Марченко, Т.В. Россада. — Київ, 2017.
2. Айвазян С. А. Прикладная статистика: Исследование зависимостей: Справ. изд. / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. — М.: Финансы и статистика, 1985.
3. Айвазян С. А. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / С. А. Айвазян, В.М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин. — М.: Финансы и статистика, 1989. .
4. Шитиков В.К., Классификация, регрессия и другие алгоритмы Data Mining с использованием R / В.К. Шитиков, С.Э. Мاستицкий. – Электронная книга, адрес доступа: <https://github.com/ranalytics/data-mining>, 2017.
5. Шлезингер М. Десять лекций по статистическому и структурному распознаванию./ М. Шлезингер, В. Главач. – Киев.: Наукова думка, 2004.
6. Макленнен Дж. Microsoft SQL Server 2008 :Datamining - интеллектуальный анализ данных. / Дж. Макленнен, Ч. Танг, Б. Криват. - СПб.: БХВ-Петербург, 2009.
7. Чубукова И.А. Data Mining: учебное пособие./ И.А. Чубукова. — М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006.
8. Leskovec J. Mining of massive datasets. / Leskovec J., Rajaraman A., Ullman J. -2019.

Додаткова

9. Пашко А.О. Аналіз динаміки ціни акцій в моделі Хестона / А.О. Пашко// Матеріали V міжнародної науково-практичної конференції "Інформаційні технології в культурі, мистецтві, освіті, науці, економіці та бізнесі". – Київ: Видавничий центр КНУКіМ, 2020. - С.136-137.
10. Барсегян А.А. Анализ данных и процессов. / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д.Тесс, С.И. Елизаров. -СПб.: БХВ-Петербург, 2009.

11. Матвеев Л.А. Компьютерная поддержка решений: Учебник. / Л.А. Матвеев. – СПб: "Специальная литература", 1998.
12. К. Карлберг. Бизнес-анализ с помощью Excel.: Пер.с англ. – К.: Диалектика, 1997. – 448с.
13. Системи підтримки прийняття рішень: Навчальний посібник/ О.І.Пушкар, В.М.Гірковатий, О.С.Євсєєв, Л.В.Потрашкова; За ред. д-ра екон. наук, проф. Пушкаря О.І.-Х.:ВД "ІНЖЕК", 2006.
14. Паклин Н. Бизнес-аналитика: от данных к знаниям. / Н. Паклин, В. Орешков. – ПИТЕР, 2010.
15. Зиновьев А.Ю. Визуализация многомерных данных./ А.Ю. Зиновьев. – Красноярск: Изд-во КГТУ, 2000.
16. Wang L. Data Mining with Computational Intelligence. / L. Wang. – Berlin: Springer-Verlag, 2005.
17. Berthold M. Intelligent Data Analysis./ M. Berthold, D.J. Hand. – Berlin: Springer-Verlag, 2007.
18. Барсегян А.А. Методы и модели анализа данных: OLAP и Data Mining. / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод/ - БХВ-Петербург, 2004.
19. Барсегян А.А. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP./ А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. - БХВ-Петербург, 2007.

10. Додаткові джерела:

https://drive.google.com/file/d/1b-zlH_h_9mO254wLoo0tJXxUPkm5zDAT/view?usp=sharing