

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

Кафедра теоретичної кібернетики

«ЗАТВЕРДЖУЮ»

Заступник декана
з навчальної роботи

_____ Кашпур О. Ф.
«__» _____ 2018 року

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
ІНТЕЛЕКТУАЛЬНА ОБРОБКА ДАНИХ**

для студентів

галузь знань **12 «Інформаційні технології»**
спеціальність **122 «Комп'ютерні науки»**
освітній рівень **бакалавр**
освітня програма **«Інформатика»**

вид дисципліни обов'язкова

Форма навчання	денна
Навчальний рік	2018/2019
Семестр	7
Кількість кредитів ECTS	4
Мова викладання, навчання та оцінювання	українська
Форма заключного контролю	екзамен

Викладачі: **д.ф.-м.н, проф. Летичевський О.А.** (лекції),

д.ф.-м.н, проф. Пашко А.О. (лекції),

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

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

КИЇВ – 2018

Розробники:

Летичевський Олександр Адольфович, член-кореспондент НАН України, д.ф.-м.н., професор, професор кафедри теоретичної кібернетики,

Пашко Анатолій Олексійович, д.ф.-м.н., доцент, професор кафедри теоретичної кібернетики

ЗАТВЕРДЖЕНО

Зав. кафедри теоретичної кібернетики

(підпис)

(Крак Ю.В.)
(прізвище та ініціали)

Протокол № ____ від «__» _____ 20__ р.

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

Протокол від «__» _____ 20__ року № ____

Голова науково-методичної комісії _____ (Хусаїнов Д.Я.)
(підпис) (прізвище та ініціали)

«__» _____ 20__ року

1 Мета дисципліни – вивчення основних та найбільш перспективних напрямків аналізу даних: зберігання інформації, оперативний і інтелектуальний аналіз. Методи та алгоритми інтелектуального аналізу.

2 Попередні вимоги до опанування або вибору навчальної дисципліни

Знати: дискретну математику, алгебру, теорію ймовірностей, математичну статистику, теорію алгоритмів та основи програмування в об'ємі стандартних університетських курсів.

Вміти: застосовувати знання з вказаних вище дисциплін до розв'язання задач.

Володіти елементарними навичками: роботи з комп'ютером

3 Анотація навчальної дисципліни

Дисципліна є вибірковою компонентою ОП підготовки фахівців за першим (бакалаврським) рівнем вищої освіти *галузі знань* 12 «Інформаційні технології» зі *спеціальності* 122 «Комп'ютерні науки», *освітньо-професійної програми* «Інформатика». Дисципліна є обов'язковою навчальною дисципліною. Викладається в 7-му семестрі, обсяг 120 год. (4 кредити ECTS), з них лекції – 54 год., консультації – 2 год., самостійна робота – 64 год. Передбачено 2 змістових модулі, 2 модульні контрольні роботи та екзамен.

В результаті вивчення навчальної дисципліни студент повинен:

знати основні методи, алгоритмічні моделі і засоби їх реалізації для здійснення інтелектуального аналізу даних;

вміти аналізувати сучасну проблематику збору та обробки інформації; використовувати моделі та методи Data Mining; використовувати методи побудови правил класифікації, дерев рішень, математичних функцій та вміння прогнозувати часові ряди;

застосовувати на практиці методи та алгоритми інтелектуальної обробки даних, комплексно використовувати аналітичні, інформаційні та комунікаційні технології при обробці та аналізі даних.

Дисципліна використовує поняття з дискретної математики, алгебри, баз даних, теорії ймовірностей та математичної статистики. Її результати використовуються в дисциплінах: «Проблеми штучного інтелекту», «Розробка програмного забезпечення», «Нейромережі та нейрообчислення».

4 Завдання (навчальні цілі)

Набуття базових знань, умінь та навичок (компетентностей) з програмування відповідно до освітньої кваліфікації «Бакалавр з комп'ютерних наук». Зокрема, розвивати:

- здатність застосовувати знання в практичних ситуаціях,
- здатність вчитися й оволодівати сучасними знаннями,
- здатність оцінювати та забезпечувати якість виконуваних робіт,
- здатність до побудови логічних висновків, використання формальних мов і моделей алгоритмічних обчислень, проектування, розроблення й аналізу алгоритмів.

5 Результати навчання за дисципліною

Результат навчання (РН) (1 – знати; 2 – вміти; 3 – комунікація; 4 – автономність та відповідальність)		Форми викладання та навчання	Методи оцінювання	Відсоток у підсумковій оцінці з дисципліни
Код	Результат навчання			
РН1.1	Знати основні методи та алгоритми, їх реалізацію сучасними мовами програмування	Лекція	Модульна контрольна робота (МКР), експерти	40
РН1.2	Знати принципи застосування основних сучасних методів та	Лекція		

	алгоритмів при розробці елементів програмного забезпечення для інтелектуальної обробки даних			
РН2.1	Вміти застосовувати сучасні алгоритми при розробці елементів програмного забезпечення	Лекція, самостійна робота	МКР, екзамен	30
РН3.1	Обґрунтовувати власний погляд на задачу, спілкуватися з колегами з питань проектування та розробки програм, складати письмові звіти	Лекція	Поточне оцінювання (ПО), екзамен	10
РН4.1	Організовувати свою самостійну роботу для досягнення результату	Самостійна робота	ПО, екзамен	10
РН4.2	Відповідально ставитися до виконуваних робіт, нести відповідальність за їх якість	Лекція	МКР, екзамен	10

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						
(з опису освітньої програми)						
ПР4. Проектувати, розробляти та аналізувати фрагменти програмного забезпечення, оцінювати їх ефективність та складність.	+	+	+	+		
ПР11. Володіти навичками використання основних сучасних алгоритмів та підходів до їх побудови при розробці елементів програмного забезпечення, сервісів інформаційних технологій відповідно до вимог і обмежень замовника, вміти готувати проектну документацію.					+	+

7 Схема формування оцінки

7.1 Форми оцінювання студентів:

- семестрове оцінювання:

Сьомий семестр

1. Модульна контрольна робота 1: РН1.1 – 30 б./9 б.

2. Модульна контрольна робота 2: РН1.1, РН2.1 – 30 б./9 б.

- підсумкове оцінювання (у формі екзамену):

- максимальна кількість балів які можуть бути отримані студентом: 40;

- результати навчання, які оцінюються: РН1.1, РН1.2, РН2.1, РН3.1;

- форма проведення: письмова

- види завдань: два теоретичні питання (40%), дві задачі (60%).

Студент допускається до екзамену, якщо в семестрі набрав не менше ніж 20 балів з кожного модуля. Для отримання загальної позитивної оцінки з дисципліни оцінка за екзамен має бути не менше 30 балів.

7.2 Організація оцінювання

Терміни проведення оцінювання в першому семестрі

1. Модульні контрольні роботи: № 1 – до 7 тижня, № 2 – до 13 тижня.

Студент має право один раз перескласти модульну контрольну роботу з можливістю отримати не більше 80% балів, призначених за роботу. Термін перескладання визначає викладач.

За відсутності студента з поважних причин перездача МКР здійснюється відповідно до «Положення про порядок оцінювання знань студентів при кредитно-модульній системі організації навчального процесу» від 1 жовтня 2010 року.

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 лекційних занять

П'ятий семестр

№ п/п	Назва теми	Кількість годин		
		Лекції	Лабора- торні	Самостійна робота
Змістовий модуль 1. Інтелектуальний аналіз даних.				
1.	Інформація і знання. Дані класифікація видів даних.	2		2
2.	Класифікація і кластеризація.	2		4
3.	Аналіз часових рядів.	4		4
4.	Дерева рішень.	2		2
5.	Методи класифікації і прогнозування.	4		4
6.	Методи кластерного аналізу.	4		4
7.	Методи пошуку асоціативних правил.	4		4
8.	Методи візуалізації.	2		4
9.	Процеси інтелектуального аналізу.	2		2
	Модульна контрольна робота 1			
Всього по модулю 1		26		30
Змістовий модуль 2. Елементи теорії груп в інтелектуальному аналізі.				
1.	Напівгрупові транзиційні системи (НТС)	6		6
2.	Еквівалентність і алгебра поведінки НТС	4		6
3.	Напівгрупове атрибутне середовище(НАС)	4		6
4.	Символьне моделювання	6		6
5.	Синтез атрибутного середовища	4		6
6.	Темпоральна логіка	4		4
	Модульна контрольна робота 2			
Всього по модулю 2		28		34

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

Лекцій – **54** год.

Самостійна робота – **64** год.

Консультації – **2** год.

Типові завдання модульних контрольних робіт

Модульна контрольна робота № 1

- Для даних відповідно до варіанту перевірте правило «Якщо ключових слів < 50%, то стаття релевантна».
- Провести ієрархічну кластеризацію одновимірного набору точок відповідно до варіанту (наприклад, 1, 4, 9, 16, 25, 36, 49, 64, 81) припускаючи, що кластери представлені своїми центроїдами і на кожному кроці об'єднуються кластери з найближчими центроїдами.

Модульна контрольна робота № 2

1. Визначити послідовну композицію умовних присвоєвань та знайти умови перестановочності.

2. Якими повинні бути δ_1 та δ_2 для того, щоб виконувалась умова $18 \leq x \leq 22$ у моделі термостату?

Питання на екзамен

1. Дайте визначення поняттю Data Mining?
2. Дайте визначення поняттю даних. Які типи даних ви знаєте?
3. Що собою являє об'єкт та атрибут в табличних даних?
4. Які типи шкал ви знаєте? В чому особливість кожної з них?
5. Які задачі, які вирішуються методами Data Mining?
6. Що собою являє Cross Industry Standard Process for Data Mining? Які його основні етапи?
7. Які методи застосовуються в Data Mining? Чим відрізняються статистичні та кібернетичні методи Data Mining?
8. Що собою являє попередній аналіз даних?
9. Які властивості методів Data Mining ви знаєте?
10. Що таке масштабованість системи?
11. В чому полягає задача класифікації даних?
12. Які типи класифікації ви знаєте?
13. Яким чином перевіряється точність класифікації?
14. Методи аналізу часових рядів.
15. Сформулюйте алгоритм побудови елементарних правил.
16. Яка закономірність лежить в основі наївного байєсівського класифікатора?
17. Про що говорить формула Баєса?
18. Сформулюйте алгоритм наївного байєсівського класифікатора.
19. Що таке активаційна функція?
20. Функції якого типу найчастіше застосовуються у якості активаційної функції і чому?
21. Сформулюйте алгоритм побудови дерева прийняття рішень?
22. Що таке ентропія і як вона застосовується при побудові дерева прийняття рішень?
23. В чому основна ідея алгоритму k-найближчих сусідів?
24. Які алгоритми KNN ви знаєте і чим вони відрізняються?
25. Сформулюйте задачу пошуку асоціативних правил.
26. Як ще називають задачу пошуку асоціативних правил?
27. Що таке асоціативне правило?
28. Як визначається достовірність та підтримка асоціативного правила?
29. Що таке узагальнені асоціативні правила?
30. Які типи асоціативних правил ви знаєте?
31. Дайте означення узагальненим та чисельним асоціативним правилам.
32. Сформулюйте задачу кластеризації
33. В чому різниця між класифікацією і кластеризацією?
34. Основні кроки алгоритму k-Means
35. Структура напівгрупової транзитивної системи (НТС).
36. Історії і траєкторії в НТС.
37. Аксиоми НТС.
38. Напівгрупа відношень переходів.
39. Напівгруповий гібридний автомат (НГА).
40. НГА є НТС.
41. Траєкторна еквівалентність НТС.
42. Бісимуляційна еквівалентність НТС.
43. Напівгрупова алгебра поведінки.
44. Локальний опис для напівгрупового атрибутного середовища (НАС).
45. Що собою являє термостат.
46. Система двох резервуарів.
47. Дворівнева система управління.
48. Символьні атрибутні середовища.
49. Предикатний трансформер.
50. Непротиворечивість і повнота визначення предикатного трансформера.

51. Формули для обчислення предикатного трансформера.
52. Редукція напівгрупових систем реального часу.
53. Мова MSC.
54. Структура моделі мови MSC.
55. Властивості історій моделі MSC.
56. Разрешающая функция моделі MSC.
57. Слабка послідовна композиція MSC-агентів.
58. Вкладені вирази мови MSC.
59. Модальна пропозиційна логіка.
60. «Філософи, що обідають».
61. Лінійна темпоральна логіка.
62. Властивості MSC-траєкторій в LTL.
63. Разрешающая функция і функция погружения MSC моделі.
64. Темпоральна логіка CTL.
65. Логіка Хенессі-Мілнера.

9. Рекомендовані джерела

Основна

1. О.О. Марченко, Т.В. Россада Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. — Київ. — 2017. — 150 с.
2. А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д.Тесс, С.И. Елизаров Анализ данных и процессов СПб.: БХВ-Петербург, 2009
3. В.В. Круглов, М.И. Дли Интеллектуальные информационные системы: компьютерная поддержка систем нечеткой логики и нечеткого вывода. — М.: Физматлит, 2002.
4. Дж. Макленнен, Ч. Танг, Б. Криват. Microsoft SQL Server 2008 :Datamining - интеллектуальный анализ данных СПб. : БХВ-Петербург, 2009.
5. И.А. Чубукова Data Mining: учебное пособие. — М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. — 382 с.
6. А.А. Летичевский, Алгебраическая теория взаимодействия и кибер-физические системы, Проблемы управления и информатика №5, 2017.
7. А.А. Летичевский, А.А. Летичевский мл., В.Г. Скобелев, В.А. Волков КИБЕР-ФИЗИЧЕСКИЕ СИСТЕМЫ, Кибернетика и системный анализ, №6, 2017.
8. Thomas A. Henzinger, The Theory of Hybrid Automata Proceedings of the 11th Annual IEEE Symposium on Logic in Computer Science (LICS 96).
9. А.А. Letichevsky, O.A. Letychevskiy, V.S. Peschanenko, T. Weigert Insertion modeling and symbolic verification of large systems, SDL Forum, Berlin.
10. A. Letichevsky and D. Gilbert. A general theory of action languages. Кибернетика и Системный Анализ, 1, 1998, 16-36, 192.
11. R. Milner, Communication and Concurrency, Prentice Hall, 1989.

Додаткова

12. Л.А. Матвеев Компьютерная поддержка решений: Учебник – СПб: "Специальная литература", 1998. – 472 с.
13. К. Карлберг. Бизнес-анализ с помощью Excel.: Пер.с англ. – К.: Диалектика, 1997. – 448с.
14. 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.-304 с.
15. Н. Паклин, В. Орешков Бизнес-аналитика: от данных к знаниям. – ПИТЕР, 2010. – 704 с.

16. С.Я. Архипенков., Д.В. Голубев, О.Б. Максименко Хранилища данных. – М.: Диалог-МИФИ, 2002.
17. А.Ю. Зиновьев Визуализация многомерных данных. – Красноярск: Изд-во КГТУ, 2000.
18. Э. Спирли Корпоративные хранилища данных. Планирование, разработка и реализация. – М.: Вильямс, 2001, т.1.
19. В.С. Рогоза, Г.В. Іщенко Інтелектуальні платформи розподілених інформаційних середовищ. – К.: НТУУ«КПІ», 2007. – 358 с.
20. В.В. Пасічник, В.А. Резніченко Організація баз даних та знань. – bhv, ПИТЕР, 2006. – 384 с.
21. L. Wang Data Mining with Computational Intelligence. – Berlin: Springer-Verlag, 2005.
22. M. Berthold, D.J. Hand (Eds.) Intelligent Data Analysis. – Berlin: Springer-Verlag, 2007.
23. А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод Методы и модели анализа данных: OLAP и Data Mining. БХВ-Петербург; 2004г.
24. А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP. БХВ-Петербург; 2007г.
25. Т. Сегаран Программируем коллективный разум. - СПб. : Символ-Плюс, 2008
26. А. Пайне Интеллектуальная обработка данных и хранилищ данных в SQL Server 2005. - М. ; СПб. : Питер, 2005
27. С.А. Айвазян, В.С Мхитарян Прикладная статистика и основы эконометрики. - М. Юнити, 1998