

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

Кафедра Комп'ютерних інтелектуальних технологій та систем



«ЗАТВЕРДЖУЮ»

Декан факультету КІУ

(Ляшенко О.С.)

(підпис, прізвище, ініціали)

«04» вересня 2019 р.

РОБОЧА ПРОГРАМА  
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

підготовки магістра

«ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ»

Галузь знань 12 Інформаційні технології  
спеціальність 123 Комп'ютерна інженерія  
за освітньою програмою «Комп'ютерні інтелектуальні технології»

Факультет Комп'ютерної інженерії та управління

Електронний документ

2019-2020 навчальний рік

Робоча програма з навчальної дисципліни **«Інтелектуальний аналіз даних»** для студентів спеціальності 123 «Комп'ютерна інженерія» за освітньою програмою «Комп'ютерні інтелектуальні технології», галузі знань 12 Інформаційні технології. [Електронний документ] - 10 с.

«02» вересня 2019 р.

Розробник: д.т.н., професор кафедри комп'ютерних інтелектуальних технологій та систем Руденко Олег Григорійович

Робоча програма затверджена на засіданні кафедри «Комп'ютерних інтелектуальних технологій та систем».

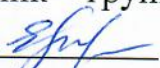
Протокол № 1 від «03» вересня 2019 р.

Завідувач кафедри КІТС

  
(підпис)

Руденко О.Г.

«03» вересня 2019 р.

Керівник групи забезпечення спеціальності 123 «Комп'ютерна інженерія»  Литвинова Є.І.  
«03» вересня 2019 р.

Схвалено методичною комісією факультету КІУ.

Протокол № 1 від «2» 09 2019 р.

Голова методичної комісії

  
(підпис)

Філіпенко І.В.

© Руденко О.Г. 2019 рік

© ХНУРЕ, 2019 рік

## 1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни	
		денна форма навчання	
Кількість кредитів ЄКТС – 4	Галузь знань 12 – « <i>Інформаційні технології</i> »	Нормативна	
	Напрямок підготовки 123 – «Комп'ютерна інженерія»		
Модулів – 1	Спеціальність:	Рік підготовки:	
Змістових модулів – 3		1-й	
Індивідуальне науково-дослідне завдання реферат: Методи інтелектуального аналізу даних в системах штучного інтелекту		Семестр	
Загальна кількість годин 120		1-й	
		Кількість годин	
		120	
		Аудиторні:	
Тижневих годин для денної форми навчання: аудиторних – самостійної роботи студента –		Освітньо-кваліфікаційний рівень: магістр	1) лекції, год 20
			2) практичні, год
			3) лабораторні, год 12
	Самостійна робота, год 88		
	в тому числі консультації, год. 8		
	в тому числі інд. завд., год. 0		
	Вид контролю: залік		

### **Примітка.**

Співвідношення кількості годин аудиторних занять до загальної кількості годин становить: 40%.

## 2 МЕТА ДИСЦИПЛІНИ ТА ОЧІКУВАНІ РЕЗУЛЬТАТИ ЇЇ ВИВЧЕННЯ

### 2.1 Мета вивчення дисципліни

Дисципліна «Інтелектуальний аналіз даних» має на меті:

- формування у студентів уявлення про типи завдань, що виникають в області інтелектуального аналізу даних (DataMining).
- вивчення основних підходів та алгоритмів розв'язання задач аналізу даних і особливостей їх застосування до вирішення реальних завдань.
- отримання студентами навичок щодо виявлення, формалізації і успішному вирішенню практичних завдань аналізу даних, що виникають в процесі їх професійної діяльності.
- отримання практичного досвіду в роботі з існуючими програмними пакетами з аналізу даних.

### 2.2 Результати вивчення дисципліни студенти повинні

**знати:** основні принципи, методи і технології інтелектуального аналізу даних; основні методи перетворення та стиснення інформації; основні методи кластеризації та прогнозування, що використовуються в аналітичних та інтелектуальних системах обробки інформації; основні методи пошуку логічних закономірностей та видобування знань з даних; принципи та особливості побудови СППР як типових систем інтелектуального аналізу даних; організацію та методи роботи зі сховищем даних як головним елементом інформаційного забезпечення систем інтелектуального аналізу даних; багатовимірні моделі даних та архітектуру OLAP-систем.

**вміти:** використовувати моделі і методи системного аналізу та математичної статистики у інтелектуальному аналізі даних (класифікація, регресія, кластеризація тощо); виконувати основні операції розподіленого інтелектуального аналізу даних; вибирати та використовувати агентні технології аналізу даних; використовувати методи та засоби Visual Mining і Text Mining; визначати можливість використання нейронних мереж та генетичних алгоритмів для інтелектуального аналізу даних.

**володіти: (перелік сформованих компетенцій)** загально-професійними, до яких відносять:

- ґрунтовна підготовка із загально-математичних дисциплін, методології системного аналізу, математичних методів розв'язання оптимізаційних задач організаційного типу, програмної та комп'ютерної інженерії;
- спеціалізовано-професійними, до яких відносять: знання сучасних теорій організації інформаційно-комп'ютерних систем, знання сучасних теорій організації баз даних та знань;
- інструментальним, до яких відносять професійне володіння комп'ютерною технікою та сучасними інформаційними технологіями.

### 2.3 Передумови для вивчення дисципліни

Вивчення курсу “Нейронні обчислювальні структури” базується на знаннях, отриманих при вивченні таких курсів, як “Вища математика”, “Теорія інформації та кодування” та “Архітектура ЕОМ”.

## 3. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1. Інтелектуальний аналіз даних: принципи, визначення, моделі, методи.

Тема 1. Місце аналізу даних в управлінні об'єктом. Призначення і мета аналізу даних. Ергономіка та естетика виробництва.

Тема 2. Основні терміни, моделі і методи інтелектуального аналізу даних (Data Mining, (DM)).

Змістовий модуль 2. Вирішення основних задач інтелектуального аналізу даних з використанням моделей і методів DM.

Тема 1. Вирішення задачі класифікації засобами DM. Методи побудови правил класифікації.

Тема 2. Вирішення задачі класифікації засобами DM. Методи побудови дерев рішень.

Тема 3. Вирішення задач регресії та аналізу часових рядків засобами DM. Метод Support Vector Machines.

Тема 4 Задача кластеризації в DM. Чіткий та нечіткий кластерний аналіз

Тема 5. Ієрархічні, неієрархічні та адаптивні методи кластеризації.

Змістовий модуль 3. Методи, засоби та інформаційні технології інтелектуального аналізу даних.

Тема 1. Візуальний аналіз даних.

Тема 2. Основні задачі аналізу текстової інформації (Text Mining).

Тема 3. Реалізація задач класифікації та кластеризації методами Text Mining.

Тема 4. Стандарти DM. Стандарт PMML.

Тема 5. Розподілений аналіз даних.

Тема 6. Принципи побудови систем розподіленого аналізу даних.

#### 4. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин					
	денна форма					
	усього	л	п	лб	конс. (ауд)	с.р.
1	2	3	4	5	6	7
<b>Модуль 1</b>						
<b>Змістовий модуль 1. Інтелектуальний аналіз даних: принципи, визначення, моделі, методи.</b>						
Тема 1. Місце аналізу даних в управлінні об'єктом. Призначення і мета аналізу даних. Ергономіка та естетика виробництва	3	1				2
Тема 2. Основні терміни, моделі і методи інтелектуального аналізу даних (Data Mining (DM))	3	1				2
Разом за змістовим модулем 1	<b>6</b>	<b>2</b>				<b>4</b>
<b>Змістовий модуль 2. Вирішення основних задач інтелектуального аналізу даних з використанням моделей і методів DM</b>						
Тема 1. Вирішення задачі класифікації засобами DM. Методи побудови правил класифікації	20	2		4		14
Тема 2. Вирішення задачі класифікації засобами DM. Методи побудови дерев рішень.	16	2				14
Тема 3. Вирішення задач регресії та аналізу часових рядків засобами DM. Метод Support Vector Machines	16	2		4		14
Тема 4. Задача кластеризації в DM. Чіткий та нечіткий кластерний аналіз	20	2				14
Тема 5. Ієрархічні, неієрархічні та адаптивні алгоритми кластеризації.	18	2		4		12
Разом за змістовим модулем 2	<b>90</b>	<b>10</b>		<b>12</b>		<b>68</b>
<b>Змістовий модуль 3. Методи, засоби та інформаційні технології інтелектуального аналізу даних</b>						
Тема 1. Візуальний аналіз даних	4	2				2
Тема 2. Основні задачі аналізу текстової інформації (Text Mining).	4	2				2

Тема 3. Реалізація задач класифікації та кластеризації методами Text Mining	4	2				2
Тема 4. Стандарти DM. Стандарт PMML.	3	1				2
Тема 5. Принципи побудови систем розподіленого аналізу даних	3	1				2
Разом за змістовим модулем 3	<b>18</b>	<b>8</b>				<b>10</b>
<b>Консультаційні заняття</b>						
Консультація № 1	2					2
Консультація № 2	2					2
Консультація № 3	2					2
Консультація № 4	2					2
Разом за інд. завд	<b>8</b>					<b>8</b>
<b>Усього годин</b>	<b>120</b>	<b>20</b>		<b>12</b>		<b>88</b>

## 5. ТЕМИ ПРАКТИЧНИХ (СЕМІНАРСЬКИХ) ЗАНЯТЬ

Семінарські заняття за навчальним планом для даної дисципліни не передбачені.

## 6 ТЕМИ ЛАБОРАТОРНИХ ЗАНЯТЬ

№ з/п	Назва теми	Кількість годин
1	Лабораторний практикум №1. Використання алгоритму k-means для вирішення задачі кластеризації.	4
2	Лабораторний практикум №2. Використання алгоритму Fuzzy C-means для вирішення задачі кластеризації.	4
3	Лабораторний практикум №3. Використання алгоритму Густафсона-Кесселя для вирішення задачі кластеризації.	4
	<b>Загальна кількість</b>	<b>12</b>

## 7 САМОСТІЙНА РОБОТА

№ з/п	Назва теми	Кількість годин	
		денна форма навчання	заочна форма навчання
1	Вивчення теоретичного матеріалу з використанням конспектів і навчальної літератури	20	
2	Підготовка до лабораторних занять	16	
3	Вивчення додаткових тем за літературними джерелами:		
	1. Методи передобробки даних	4	
	2. Зниження розмірності та компресія даних. Аналіз головних компонент, факторний аналіз	4	
	3. Тест FASMI	4	
	4. Алгоритм C4.5 як метод побудови дерева рішень	4	
	5. Алгоритм покриття як метод побудови дерева рішень	4	
	6. Приклад роботи алгоритму Apriori.	4	
	7. Алгоритм k-means	4	
	8. Алгоритм Fuzzy C-means	4	
	9. Алгоритм Густавсона-Кесселя	2	
	10. Вилучення ключових понять з тексту	2	

11. Анотування текстів методами Text Mining	4	
12. Засоби аналізу текстової інформації	4	
13. Стандарт JDMAPI	4	
14. Основи JADE	4	
15. Основи нейронних мереж	4	
16. Основи побудови генетичних алгоритмів	4	
Загальна кількість, год	<b>88</b>	

## 8 ІНДИВІДУАЛЬНІ ЗАВДАННЯ

Індивідуальні заняття за навчальним планом для даної дисципліни відсутні.

## 9. МЕТОДИ НАВЧАННЯ ТА ЗАСОБИ ОЦІНЮВАННЯ

Лекції, лабораторні роботи, самостійна робота.

## 10 МЕТОДИ КОНТРОЛЮ ТА РЕЙТИНГОВА ЗА ДИСЦИПЛІНОЮ

### 10.1 Розподіл балів, які отримують студенти (кількісні критерії оцінювання)

Вид заняття / контрольний захід	Оцінка
Лб № 1, 2,	$(3...5) \times 2 = 6...10$
ДКР1	24-40
<b>Контрольна точка I</b>	<b>30-50</b>
Лб № 3	$(3...5) \times 1 = 3...5$
ДКР2	13-20
ТЕСТ	14-20
<b>Контрольна точка II</b>	<b>30-50</b>
<b>Всього за семестр</b>	<b>60-100</b>

Як форма підсумкового контролю для дисципліни «Нейронні обчислювальні структури» використовується письмовий (комбінований) іспит. При цьому виді контролю підсумкова оцінка  $P_{\Pi}$  обчислюється за формулою:  $P_{\Pi} = 0,6 \cdot O_{\text{сем}} + 0,4 \cdot O_{\text{ісп}}$ , де  $O_{\text{сем}}$  – оцінка за семестр у 100-бальній системі,  $O_{\text{ісп}}$  – оцінка за іспит у 100-бальній системі.

Для оцінювання роботи студента протягом семестру підсумкова рейтингова оцінка  $O_{\text{сем}}$  розраховується як сума оцінок за різні види занять та контрольні заходи. Кожна лабораторна робота оцінюється в 5 балів (1 бал за присутність + 1 бал за відпрацювання + 3 бали за захист (задача з оцінкою)). ДКР1 оцінюються в 21-35 балів, ДКР2 – в 12-20 балів, Тест – в 12-20 балів. Максимальна рейтингова оцінка протягом семестру – 100 балів.

### 10.2 Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки

1. Визначення терміну «аналіз». Місце аналізу в управлінні об'єктами. Мета і призначення аналізу даних. Тенденції розвитку баз даних. Визначення терміну «сховище даних». Види сховищ даних, їх переваги та недоліки. Визначення терміну «OLAP». Підходи до створення та особливості OLAP-систем. Визначення терміну «Інтелектуальний аналіз даних» («Data Mining»). Основні моделі і методи інтелектуального аналізу даних. Способи

представлення даних, їх передобробки та компресії. Галузі використання інтелектуального аналізу даних.

2. Визначення терміну «класифікація» в інтелектуальному аналізі даних. Формальна постановка задачі класифікації. Методи вирішення задачі класифікації. Засоби представлення результатів вирішення задачі класифікації. Класифікаційні правила. Метод Naive Bayes. Деревя рішень та можливість їх перетворення на класифікаційні правила. Проблеми формування дерев рішень. Алгоритм ID3. Алгоритм C4.5. Алгоритм покриття. Формальна постановка задачі регресії. Формальна постановка задачі аналізу часових рядів. Метод найменших квадратів. Метод Support Vector Machines. Формальна постановка задачі пошуку асоціативних правил. Різновиди асоціативних правил. Секвенційний аналіз. Алгоритм Apriori. Формальна постановка задачі кластеризації. Метрики та міри відстані між кластерами і об'єктами. Представлення результатів вирішення задачі кластеризації. Ієрархічні алгоритми вирішення задачі кластеризації. Дивизимні алгоритми. Агломеративні алгоритми. Неієрархічні алгоритми вирішення задачі кластеризації. Алгоритм k-means. Алгоритм Fuzzy C-Means.

3. Визначення терміну «візуальний аналіз даних». Особливості та етапи виконання візуального аналізу даних. Характеристики засобів візуального аналізу даних. Основні методи виконання візуального аналізу даних. Визначення терміну «Text Mining». Етапи Text Mining. Прийоми попередньої обробки текстів. Типові задачі Text Mining. Вилучення ключових понять з тексту. Задача класифікації у Text Mining. Кластеризація текстових документів. Напрямки стандартизації інтелектуального аналізу даних. Виникнення стандарту CRISP. Структура стандарту CRISP. Фази і задачі інтелектуального аналізу даних за стандартом CRISP. Призначення стандарту CWM. Структура стандарту CWM. Використання стандарту CWM в типових технологіях інтелектуального аналізу даних. Призначення стандарту PMML. Структура стандарту PMML. Особливості використання стандарту PMML. Особливості проведення розподіленого аналізу даних. Методи і технології розподіленого аналізу даних. Основні особливості еволюційних алгоритмів аналізу даних. Основні особливості агентних технологій розподіленого аналізу даних. Принципи побудови систем і технологій розподіленого аналізу даних.

### **Критерії оцінювання роботи студента протягом семестру**

*Задовільно, D, E (60-74).* Мати мінімум знань та умінь. Відпрацювати та захистити всі лабораторні практикуми. Вміти використовувати основні методи інтелектуального аналізу даних.

*Добре, C (75-89).* Твердо знати мінімум теоретичних знань. Показати вміння виконувати та захищати всі лабораторні практикуми в обумовлений викладачем термін з обґрунтуванням рішень та заходів, які запропоновано у роботах.

*Відмінно, A, B (90-100).* Показати повні знання основного та додаткового теоретичного матеріалу. Безпомилково виконувати та захищати всі лабораторні практикуми в обумовлений викладачем термін з докладним обґрунтуванням рішень та заходів, які запропоновано у роботах.

### **Шкала оцінювання: національна та ECTS**

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	<b>A</b>	відмінно	зараховано
82-89	<b>B</b>	добре	
74-81	<b>C</b>		
64-73	<b>D</b>	задовільно	
60-63	<b>E</b>		
35-59	<b>FX</b>	незадовільно з можливістю	не зараховано з



		повторного складання	можливістю повторного складання
0-34	<b>F</b>	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

## 11 МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ТА РЕКОМЕНДОВАНА ЛІТЕРАТУРА

### 11.1 Базова література

1. Гудфеллоу Я. Глубокое обучение/ Я. Гудфеллоу, И. Бенджио, А. Курвилль / пер. с англ. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с
2. Николенко С. Глубокое обучение. / С. Николенко, А. Кадурич, Е. Архангельская — СПб.: Питер, 2018. — 480 с.
3. Бринк Х. Машинное обучение. / Х. Бринк, Д. Ричардс, М. Феверолф -СПб.: Питер, 2017. -336 с.
4. Шакла Н. Машинное обучение и TensorFlow. / Н. Шакла - СПб.: Питер, 2019. - 336 с.
5. Гифт Н. Прагматичный ИИ. Машинное обучение и облачные технологии. / Н. Гифт - СПб.: Питер, 2019. - 304 с.
6. Чубукова, И.А. Data Mining / И.А. Чубукова. - 2-е изд., испр. - М. :Интернет-Университет Информационных Технологий, 2008. - 383 с. -(Основы информационных технологий). - ISBN 978-5-94774-819-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=233055>
- 2 Шумейко А. А. Интеллектуальный анализ данных (Введение в Data Mining) : учеб. пособ. / А. А. Шумейко, С. Л. Сотник. – Днепропетровск: Белая Е. А., 2012. – 212 с.
- 3 Луньков А.Д. Интеллектуальный анализ данных учеб. пособ. / А.Д. Луньков, А.В. Харламов – Саратов, 2008. – 96 с.
- 4 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. – СПб.: БХВ-Санкт-Петербург, 2008. – 384 с.
- 5 Aggarwal С.С. Data Mining The Textbook / С.С. Aggarwal. – New York: Springer-Verlag, 2015. – 746 p.
- 6 Han, J. Data Mining: Concepts and Techniques / J. Han, M. Kamber. – Amsterdam e.a.: Morgan Kaufmann Publishers, 2006. – 754 p.
- 7 Olson, D.L. Advanced Data Mining Techniques / D.L. Olson, D. Dursun. – Berlin: Springer, 2008. – 180 p.
- 8 Maimon, O. Soft Computing for Knowledge Discovery and Data Mining / O. Maimon, L. Rokach. – New York: Springer-Verlag, 2007. – 448 p.
- 9 Xu R. Clustering / R. Xu, D.C. Wunsch. – Hoboken, New Jersey: John Wiley & Sons, Inc., 2009. – 341 p.
- 10 Дюк В. Data Mining / В. Дюк, А. Самойленко. – Санкт-Петербург: Питер, 2001. – 368 с.

### 11.2 Допоміжна література

- 11 Aggarwal С.С. Data Streams: Models and Algorithms/ С.С. Aggarwal. – New York: Springer-Verlag, 2007. – 373 p.
- 12 Браверман, Э.М. Структурные методы обработки эмпирических данных / Э.М. Браверман, И.Б. Мучник. – Москва: Наука, 1983. – 464 с.

13 Айвазян, С.А. Прикладная статистика. Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – Москва: Финансы и статистика, 1989. – 607 с.

## 12 ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ

1. Конспект лекцій з дисципліни “Інтелектуальний аналіз даних” для студентів денної форми навчання за спеціальністю 8.080401 – “Інформаційні управляючі системи і технології” / Упоряд. М.В. Євланов (електронний варіант). – 90 с.

2. Методичні вказівки до самостійної роботи з дисципліни “Інтелектуальний аналіз даних” для студентів денної форми навчання за спеціальністю 8.080401 – “Інформаційні управляючі системи і технології” / Упоряд. М.В. Євланов (електронний варіант). – 63 с.

3. Методичні вказівки до лабораторних робіт з дисципліни “Інтелектуальний аналіз даних” для студентів денної форми навчання за спеціальністю 8.080401 – “Інформаційні управляючі системи і технології” / Упорядники В.М. Левикін, М.В. Євланов, М.А. Керносов (електронний варіант). – 28 с.

4. Бодянский, Е.В. Гибридные нейро-фаззи модели и мультиагентные технологии в сложных системах / Е.В. Бодянский, В.Е. Кучеренко, Е.И. Кучеренко, А.И. Михалев, В.А. Филатов. – Днепропетровск: Системные технологии, 2008. – 403 с.

5 <http://www.machinelearning.ru/> колекція матеріалів по машинному обучению

6 <http://archive.ics.uci.edu/ml/> колекція прикладних задач 7 Програмне забезпечення з дисципліни «Інтелектуальний аналіз даних»