

Міністерство освіти і науки України  
Запорізький національний технічний університет

**МЕТОДИЧНІ ВКАЗІВКИ**  
до виконання самостійної роботи  
з дисципліни  
**"ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ"**  
для студентів  
напряму підготовки 6.050101 "Комп'ютерні науки"  
всіх форм навчання

2017

Методичні вказівки до виконання самостійної роботи з дисципліни «Інтелектуальний аналіз даних» для студентів напряму підготовки 6.050101 «Комп'ютерні науки» всіх форм навчання / Укл.: Т.В. Федорончак. – Запоріжжя: ЗНТУ, 2017. – 17 с.

Автори: Т.В. Федорончак, к.т.н., доцент

Рецензент: В.І. Дубровін, к.т.н., професор

Відповідальний  
за випуск: С.О. Субботін, д.т.н., професор

Затверджено  
на засіданні кафедри  
програмних засобів

Протокол №11  
від "06" червня 2017 р.

**ЗМІСТ**

Вступ.....	4
1. Мета і завдання дисципліни, її місце в навчальному процесі.....	5
2. Робоча програма дисципліни.....	7
3. Контрольні заходи з перевірки якості засвоєння навчального матеріалу дисципліни.....	15
4. Рекомендована література .....	16

## ВСТУП

Метою вивчення дисципліни «Інтелектуальний аналіз даних» є вивчення методів сучасної обробки даних – аналітичного дослідження великих масивів інформації з метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів і різних інструментальних засобів, які використовуються в аналізі даних; розгляд практичних прикладів застосування інтелектуального аналізу даних; підготовка студентів до самостійної роботи з вирішення задач засобами інтелектуального аналізу даних і розробки інтелектуальних систем.

Термін, що передбачений робочим планом на аудиторні заняття з дисципліни, не дає можливості у необхідному обсязі викласти навчальний матеріал дисципліни. передбачається самостійне опанування студентом частини курсу. До того ж, матеріал, що викладається в аудиторії, теж повинен бути закріплений шляхом самостійної роботи студента.

Мета виконання самостійної роботи – поглиблення, узагальнення і закріплення теоретичних знань і практичних умінь студентів з дисципліни, що вивчається, шляхом вироблення вміння самостійної роботи з навчальною і фаховою науково-технічною літературою.

Мета методичних рекомендацій полягає в наступному:

- ознайомити студентів з повним обсягом навчального матеріалу з дисципліни, який він повинен засвоїти, в тому числі і з тією частиною, яка повністю виноситься на самостійне вивчення;
- навести необхідну навчальну літературу по кожній тематиці дисципліни;
- надати методичні вказівки та контрольні питання для самоперевірки знань;
- ознайомити студентів з заходами контролю засвоєння навчального матеріалу з дисципліни.

# 1. МЕТА І ЗАВДАННЯ ДИСЦИПЛІНИ, ЇЇ МІСЦЕ В НАВЧАЛЬНОМУ ПРОЦЕСІ

## 1.1 Мета вивчення дисципліни

**Мета** – вивчення методів сучасної обробки даних – інтелектуального аналізу даних (Data Mining, Knowledge Discovery in Data), аналітичного дослідження великих масивів інформації з метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів і різних інструментальних засобів, які використовуються Data Mining; розгляд практичних прикладів застосування Data Mining; підготовка студентів до самостійної роботи з вирішення задач засобами Data Mining і розробки інтелектуальних систем.

Data Mining – мультидисциплінарна область, яка виникла і розвивається на базі таких наук як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних та ін.

## 1.2 Завдання вивчення дисципліни

**Завдання** дисципліни полягає в тому, що студенти повинні :

- опанувати базові принципи побудови моделей даних;
- ознайомитися з концепцією Knowledge Discovery in Data (виявлення знань в даних) и Data Mining («видобування» знань);
- навчитися ефективно використовувати методи здобуття знань з великих масивів даних;
- ознайомитися з основними типами задач, що можуть бути вирішені за допомогою методів інтелектуального аналізу даних;
- отримати практичні навички з використання інструментальних засобів інтелектуального аналізу даних при вирішенні прикладних задач та навчитися інтерпретувати отримані результати.

У результаті вивчення навчальної дисципліни студент повинен

**знати :**

- основні поняття, задачі та стадії інтелектуального аналізу даних;
- підходи к збереженню, представленню та обробці інформації в сучасних інформаційних системах;

- методи побудови моделей та аналізу залежностей у великих масивах даних;

- сучасні програмні засоби для проектування і розробки систем інтелектуального аналізу даних;

- концепції сховищ даних, їх оперативної аналітичної обробки;

**В М І Т И :**

- обґрунтовувати вибір конкретного типу моделі та методу інтелектуального аналізу даних при вирішенні поставленої практичної задачі;

- проводити необхідну попередню обробку даних, визначати тип задачі аналізу, вирішувати її адекватно обраним методом з оптимально визначеними параметрами, оцінювати результати, робити змістовні висновки та інтерпретацію;

- використовувати сучасні програмні засоби для проектування та дослідження систем інтелектуального аналізу даних;

- застосовувати технології роботи зі сховищами даних, здійснювати їх аналітичну обробку та інтелектуальний аналіз для забезпечення надійної роботи інформаційних систем;

- проектувати інформаційне забезпечення (логічну та фізичну структури баз даних) інформаційних систем.

### **1.3 Перелік дисциплін, засвоєння яких є необхідним для вивчення дисципліни**

Курс базується на знаннях, одержаних при вивченні дисциплін «Вища математика», «Організація баз даних та знань», «Математичні методи дослідження операцій» та «Теорія ймовірностей та математична статистика».

Отриманні знання будуть використовуватися та доповнюватися при подальшому вивченні дисциплін «Теорія прийняття рішень», «Методи та системи штучного інтелекту».

## 2. РОБОЧА ПРОГРАМА ДИСЦИПЛІНИ

Наведено зміст дисципліни «Інтелектуальний аналіз даних», її структуру та кількість годин на лекційні, лабораторні заняття, на самостійну роботу студента при вивченні кожної теми.

Згідно діючого навчального плану викладання дисципліни «Інтелектуальний аналіз даних» здійснюється у 5 семестрі програми підготовки бакалаврів загальною кількістю 105 годин (3,5 кредити), в тому числі лекції – 28 години, лабораторні заняття – 14 години, індивідуальна самостійна робота студента – 63 годин. По закінченню семестру передбачено залік з дисципліни.

### 2.1 Зміст дисципліни

**Змістовий модуль 1. Основи інтелектуального аналізу даних: процес виявлення знань**

**Тема 1.** Основи інтелектуального аналізу даних.

Визначення Data Mining і область застосування. Задачі, моделі та методи Data Mining. Методи, стадії, задачі Data Mining. Поняття Business Intelligence.

Література [3-6, 8-9].

**Тема 2.** Процес виявлення знань.

Цикл одержання, попередньої обробки, аналізу даних, інтерпретації результатів та їхнього використання. Етапи процесу Data Mining, пов'язані з побудовою, перевіркою, оцінкою, вибором і корекцією моделей. Методи первісної обробки даних. Інструментальні засоби Data Mining. Методи дослідження структури даних: візуалізація даних.

Література [3-6, 8-9].

**Змістовий модуль 2. Алгоритми Data Mining: класифікація і регресія**

**Тема 1.** Вирішення задачі класифікації

Постановка задачі класифікації та представлення результатів. Методи побудови правил класифікації. Методи побудови дерев рішень. Методи побудови математичних функцій. Методи опорних векторів, «найближчого сусіда», Байеса. Аналіз багатомірних

угруповань. Класифікація об'єктів у випадку невідомих розподілень даних. Методи оцінювання помилок класифікації.

Література [3-6, 8-9].

### **Тема 2. Вирішення задачі регресії**

Сутність задачі прогнозування. Методи вирішення задачі регресії.

Література [3-6, 8-9].

### **Змістовий модуль 3. Інтелектуальний аналіз часових рядів**

#### **Тема 1. Методи аналізу часових рядів**

Поняття нечітких часових рядів. Методи моделювання часових рядів. Методи аналізу та прогнозування поведінки часових рядів.

Література [3-6, 8-9].

### **Змістовий модуль 4. Алгоритми Data Mining: кластеризація**

#### **Тема 1. Вирішення задачі кластеризації**

Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.

Література [3-6, 8-9].

### **Змістовий модуль 5. Алгоритми Data Mining: пошук асоціативних правил**

#### **Тема 1. Вирішення задачі пошуку асоціативних правил**

Постановка задачі пошуку асоціативних правил та представлення результатів. Секвенціальний аналіз. Різновиди задач пошуку асоціативних правил. Методи подання результатів. Алгоритми пошуку асоціативних правил. Методи пошуку асоціативних правил: метод Apriori, побудова FP-дерев пошуку шаблонів даних. Min-max асоціації у базах даних. Побудова hash-дерев.

Література [3-6, 8-9].

### **Змістовий модуль 6. Сховища даних та оперативний аналіз даних (OLAP)**

#### **Тема 1. Сховища даних**

Визначення сховища даних, порівняння з базами даних, використання. Архітектура сховища даних. ETL-процеси (добування,



перетворення й завантаження даних). Вітрини даних, куби даних, багатомірна модель даних.

Література [3-6, 8-9].

### **Тема 2. Оперативний аналіз даних**

Розгортання OLAP-кубів. Операції над OLAP-кубами (зріз, обертання, консолідація, деталізація). Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.

Література [3-6, 8-9].

## **2.2 Структура дисципліни**

Назви змістових модулів і тем	Кількість годин					
	у тому числі					
	усього	л	п	лаб	інд	с.р.
<b>Змістовий модуль 1. Основи інтелектуального аналізу даних, процес виявлення знань</b>						
Тема 1. Основи інтелектуального аналізу даних	5	2	-	-	-	3
Тема 2. Процес виявлення знань	10	2	-	2	-	6
Разом за змістовим модулем 1	15	4	-	2	-	9
<b>Змістовий модуль 2. Алгоритми Data Mining: класифікація і регресія</b>						
Тема 1. Вирішення задачі класифікації	15	6	-	4	-	5
Тема 2. Вирішення задачі регресії	10	2	-	2	-	6
Разом за змістовим модулем 2	25	8	-	6	-	11
<b>Змістовий модуль 3. Інтелектуальний аналіз часових рядів</b>						
Тема 1. Методи аналізу часових рядів	10	2	-	-	-	8
Разом за змістовим модулем 3	10	2	-	-	-	8
<b>Усього годин</b>	<b>50</b>	<b>14</b>	<b>-</b>	<b>8</b>	<b>-</b>	<b>28</b>

Назви змістових модулів і тем	Кількість годин					
	у тому числі					
	усього	л	п	лаб	інд	с.р.
<b>Змістовий модуль 4. Алгоритми Data Mining: кластеризація</b>						
Тема 1. Вирішення задачі кластеризації	15	6	-	4		5
Разом за змістовим модулем 4	15	6	-	4	-	5
<b>Змістовий модуль 5. Алгоритми Data Mining: пошук асоціативних правил</b>						
Тема 1. Вирішення задачі пошуку асоціативних правил	15	4	-	2	-	9
Разом за змістовим модулем 5	15	4	-	2		9
<b>Змістовий модуль 6. Сховища даних та оперативний аналіз даних (OLAP)</b>						
Тема 1. Сховища даних	10	2	-	-	-	8
Тема 2. Оперативний аналіз даних	15	2	-	-	-	13
Разом за змістовим модулем 5	25	4	-	-		21
<b>Усього годин</b>	55	14	-	6	-	35
<b>Разом</b>	105	28	-	14	-	63

### 2.3 Перелік лабораторних занять та їх тривалість

Метою практичної частини курсу «Інтелектуальний аналіз даних» є отримання практичних навичок використання розглянутих на лекціях методів аналізу для вирішення практичних завдань.

WEKA (Waikato Environment for Knowledge Analysis) – бібліотека алгоритмів машинного навчання для вирішення завдань інтелектуального аналізу даних. Система дозволяє безпосередньо застосовувати алгоритми до вибірок даних, а також викликати алгоритми з програм на мові Java.

Користувачами WEKA є дослідники в області машинного

навчання і прикладних наук. Вона також широко використовується в навчальних цілях.

Теоретичні відомості, присвячені використуванним алгоритмам інтелектуального аналізу даних, можна отримати з конспекту лекцій з дисципліни «Інтелектуальний аналіз даних» або з рекомендованих літературних джерел. При виконанні лабораторних робіт слід розібратися з кодом реалізації розглянутих алгоритмів у програмі WEKA.

Загальний залік з практичної частини курсу студент одержує після виконання і здачі всіх лабораторних робіт.

Для виконання передбачених лабораторних робіт використовуються розроблені методичні вказівки [2].

**Лабораторна робота №1.** Знайомство з програмою інтелектуального аналізу даних WEKA та підготовка даних (2 години)

Мета роботи - ознайомитися та отримати навички роботи з GUI інтерфейсом бібліотеки data mining алгоритмів WEKA, вивчити можливості, що надаються програмою WEKA; на практиці вивчити методи попередньої обробки даних для виконання задач інтелектуального аналізу даних.

**Лабораторна робота №2.** Задача класифікації (4 години)

Мета роботи – на практиці вивчити роботу алгоритмів класифікації, навчитися інтерпретувати результати роботи класифікаторів і вибирати найкращий метод для вирішення поставленої задачі.

**Лабораторна робота №3.** Прогнозування, задача регресії (2 години)

Мета роботи – на практиці вивчити роботу алгоритмів, що вирішують задачу регресії, і навчитися інтерпретувати результати їх роботи.

**Лабораторна робота №4.** Задача кластеризації (4 години)

Мета роботи – на практиці вивчити роботу алгоритмів кластеризації, навчитися інтерпретувати результати їх роботи і вибирати найкращий метод для розв'язуваної прикладної задачі.

**Лабораторна робота №5.** Пошук асоціативних правил (2 години)

Мета роботи – на практиці вивчити роботу алгоритмів пошуку асоціативних правил та навчитися інтерпретувати результати їх роботи.

## **2.4 Зміст самостійної роботи**

Самостійна робота студента: (денна форма навчання) опрацювання матеріалу, прослуханого на лекціях, самостійне опанування частини теоретичного матеріалу, робота з контрольними запитаннями та завданнями, підготовка до лабораторних робіт, ознайомлення з програмним забезпеченням, що використовується в практичній частині курсу, опрацювання навчальної та методичної літератури, підготовка до заліку – 63 год.

Студенти заочної форми навчання також виконують контрольну роботу.

На самостійну роботу студентів значною мірою виносяться наступні теми:

1. Сховища даних (8 годин);
2. Оперативний аналіз даних (13 годин);
3. Прогнозування часових рядів (8 годин).

## **2.5 Контрольні питання**

При підготовці до поточного та остаточного контролю знань студент може перевірити свою готовність, відповідаючи на нижченаведені питання, які охоплюють вузлові положення дисципліни «Інтелектуальний аналіз даних».

1. Дайте визначення інтелектуального аналізу даних.
2. Що таке розвідувальний аналіз.
3. Що таке генеральна сукупність і вибірка?
4. Які властивості повинна мати вибірка?
5. Що таке якість даних?
6. Які цілі підготовки даних до аналізу? Які завдання в неї входять?
7. Який атрибут називається цільовим?

8. Що таке значимий і незначний атрибут?
9. Що таке відбір атрибутів?
10. У чому полягає завдання класифікації? Наведіть практичний приклад.
11. Що таке навчання з учителем і без учителя? До якого типу відноситься задача класифікації?
12. Завдання класифікації є описовим або прогнозуючим і чому?
13. Навіщо потрібна навчальна і тестова вибірки для вирішення завдання класифікації?
14. Які існують підходи для поділу вихідної вибірки на навчальну і тестову?
15. Метод класифікації 0R.
16. Метод класифікації за одним правилом 1R.
17. Метод класифікації PRISM.
18. Наївна Баєсова класифікація NaiveBayes.
19. Метод побудови дерев рішень CART
20. Метод побудови дерев рішень Id3 та C4.5.
21. Метод опорних векторів SMO.
22. Метод k найближчих сусідів kNN.
23. Як оцінити якість побудованої моделі класифікації?
24. Що таке матриця помилок? Як її інтерпретувати?
25. Що означають параметри чутливість, специфічність, точність? Як їх розрахувати?
26. Що таке параметр Каппа? Що він показує?
27. Що таке аналіз витрати-вигоди?
28. Як порівняти роботу двох класифікаторів?
29. Що таке ансамблі класифікаторів і адаптивний бустінг? Для чого вони застосовуються?
30. У чому полягає завдання регресії? Наведіть приклад.
31. Чим завдання регресії схоже і чим відрізняється від завдання класифікації?
32. Що таке навчання з учителем і без учителя? До якого типу належить завдання регресії?
33. Завдання регресії є описовим або прогнозуючим і чому?
34. Метод лінійної регресії.
35. Метод опорних векторів для вирішення завдання регресії.
36. Метод найближчих сусідів для вирішення завдання регресії.

37. Вирішуючі дерева для завдання регресії.
38. Як оцінити якість побудованої моделі для завдання регресії?
39. Прогнозування часових рядів.
40. У чому полягає завдання кластеризації? Наведіть практичний приклад.
41. Що таке навчання з учителем і без учителя? До якого типу відноситься завдання кластеризації?
42. Завдання кластеризації є описовим або прогнозуючим і чому?
43. Чим визначається «схожість» об'єктів при вирішенні задачі кластеризації?
44. Що таке однорівнева і ієрархічна кластеризація?
45. Що таке чітка і нечітка кластеризація?
46. Які є підходи до розрахунку відстані між кластерами?
47. Алгомеративная і дівізімная ієрархічна кластеризація?
48. Метод кластеризації k-середніх.
49. Метод нечіткої кластеризації fuzzy c-means.
50. Метод кластеризації k-медоїд.
51. Ієрархічні методи кластеризації.
52. Імовірнісний метод кластеризації EM/
53. Метод кластеризації COBWEB.
54. Метод кластеризації заснований на щільності розташування об'єктів DBSCAN.
55. Як оцінити якість побудованої моделі для завдання кластеризації?
56. У чому полягає завдання пошуку асоціативних правил? Наведіть практичний приклад.
57. Що таке частий набір?
58. Що таке сильне асоціативне правило?
59. З яких двох кроків складається пошук асоціативних правил?
60. У чому полягає принцип Apriori?
61. Як формуються правила зі знайдених частих наборів?
62. Опишіть алгоритм Apriori
63. Що означають параметри support, confidence, lift, conviction, що застосовуються в алгоритмі Apriori?
64. Опишіть алгоритм ECLAT
65. Опишіть алгоритм FPGrowth

### **3. КОНТРОЛЬНІ ЗАХОДИ З ПЕРЕВІРКИ ЯКОСТІ ЗАСВОЄННЯ НАВЧАЛЬНОГО МАТЕРІАЛУ ДИСЦИПЛІНИ**

Формами контролю студентів за якістю оволодіння навчальним матеріалом є: самоконтроль за допомогою контрольних-тестових завдань та контроль з боку викладача, який здійснюється за допомогою методів поточного і підсумкового контролю.

Керівництво самостійною роботою студентів – це індивідуально-консультативна робота, що проводиться з метою посилення мотивації студентів до пізнавальної діяльності і спрямування її в необхідному напрямку.

Індивідуально-консультативна робота, як правило, проводиться у вигляді консультацій, інколи – індивідуальних занять у формі: діалогу з різних навчальних проблем; перевірки виконання завдань; виконання індивідуальних завдань (курсівих та дипломних проєктів (робіт), розрахункових, творчих робіт тощо); евристичної бесіди; наукової роботи та ін.

Контрольні заходи з дисципліни «Інтелектуальний аналіз даних» передбачають наступне.

Контроль успішності студентів денної форми навчання здійснюється за результатами:

- усне або письмове опитування з кожної теми;
- захисту звітів про виконання лабораторних робіт;
- заліку в кінці семестру (за умови виконання студентом вимог навчального процесу).

Контроль успішності студентів заочної форми навчання здійснюється за результатами:

- захисту звітів про виконання лабораторних робіт;
- захисту контрольної роботи;
- заліку в кінці семестру.

Склад, обсяг і терміни виконання змістових модулів, на які підрозділяється робоча програма дисципліни, надано у розділі 2.

Підсумкова оцінка за дисципліну розраховується за результатом захисту лабораторних робіт та результатів проміжних опитувань.

## 4. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Надається перелік навчальної та довідникової літератури, що рекомендується при вивченні дисципліни «Інтелектуальний аналіз даних». Слід мати на увазі, що джерела, які можуть бути використані, не обмежуються тільки цим переліком.

### 4.1 Навчально-методичне забезпечення

1. Комплект слайдів презентацій з матеріалами лекцій.
2. Методичні вказівки до лабораторних робіт з дисципліни «Інтелектуальний аналіз даних» для студентів напрямів підготовки 6.050103 «Програмна інженерія» та 6.050101 «Комп'ютерні науки» / Т.В. Федорончак. – Запоріжжя: ЗНТУ, 2016. – 62 с.

### 4.2 Базова література

3. Барсегян А. А. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.
4. Олійник А. О. Інтелектуальний аналіз даних : Навчальний посібник / А. О. Олійник, О. О. Олійник, С. О. Субботін. – Запоріжжя : ЗНТУ, 2011. – 278 с.
5. Ситник В.Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): навч. посібник. – К.: КНЕУ, 2007. – 376 с.
6. Чубукова И.А. Data Mining: учебное пособие / И.А. Чубукова. – М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. – 382 с.
7. Дубровин В.И. Интеллектуальные средства диагностики и прогнозирования надежности авиадвигателей: Монография / В.И. Дубровин, С.А. Субботин, А.В. Богуслаев, В.К. Яценко. – Запорожье: ОАО "Мотор-Сич", 2003. – 279 с.
8. Han J. Data Mining: Concepts and Techniques (Second Edition) / J. Han, M. Kamber – Morgan Kaufmann Publishers, 2006. – 800 p.
9. Witten, I. H. Data mining : practical machine learning tools and techniques. / Ian H. Witten, Frank Eibe, Mark A. Hall. – 3rd ed. – Morgan Kaufmann Publishers, 2011. – 630 p.



### 4.3 Допоміжна література

10. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям / Н.Б. Паклин, В.И. Орешков. – СПб.: Питер, 2009. – 624 с.
11. Сегаран Т. Программируем коллективный разум / Т. Сегаран. – СПб.: Символ-Плюс, 2008. – 368 с.
12. Макленнен Д. Microsoft SQL Server 2008: Data Mining – интеллектуальный анализ данных / Д. Макленнен, Ч. Танг, Б. Криват. – СПб.: БХВ-Петербург, 2009. – 720 с.
13. 9. Дюк В. Data Mining : учебный курс / В. Дюк, А. Самойленко. – СПб.: Питер, 2001. – 368 с.

### 4.4 Інформаційні ресурси

14. Weka 3: Data Mining Software in Java [Електронний ресурс] – Режим доступу: <http://www.cs.waikato.ac.nz/ml/weka/>
15. Weka 3 Wiki documentation [Електронний ресурс] – Режим доступу: <http://weka.wikispaces.com/>