

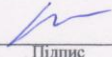


**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

**Аналітика великих даних**


**Статус дисципліни:** вибіркова дисципліна  
**Факультет** – Інформаційних технологій  
**Кафедра** – Комп'ютерної інженерії та інформаційних систем

Форма навчання	Курс	Семестр	Загальне навантаження		Кількість годин						Форма семестрового контролю			
			Кредити ЕКТС	Години	Аудиторні заняття				Індивідуальна робота студента	Самостійна робота, в т.ч. ІРС	Курсовий проект	Курсова робота	Залік	Іспит
					Разом	Лекції	Лабораторні роботи	Практичні заняття						
Д		непарний	8	240	85	34	34	17		155			+	
<b>Разом ДФН</b>			<b>8</b>	<b>240</b>	<b>85</b>	<b>34</b>	<b>34</b>	<b>17</b>		<b>155</b>			<b>1</b>	

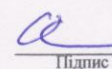
Програма складена  Т. М. Кисіль  
 Підпис Ініціали, прізвище викладача(ів)

Схвалена на засіданні кафедри комп'ютерної інженерії та інформаційних систем

Протокол № 1 від 12 серпня 2022 р.

Зав. кафедри комп'ютерної інженерії та інформаційних систем  Т. О. Говорущенко  
 Підпис Ініціали, прізвище

Робоча програма розглянута та схвалена Вченою радою факультету інформаційних технологій

Голова Вченої ради  Савенко О.С.  
 Підпис Ініціали, прізвище

## ПОЯСНЮВАЛЬНА ЗАПИСКА

Дисципліна «Аналітика великих даних» є дисципліною прикладного спрямування, яка покликана закріпити та розвинути в здобувачів ступеня магістра навичок використання сучасних методів аналізу великих даних.

**Мета дисципліни:** сформувати у студентів систему знань, умінь та навичок застосування сучасних методів та технологій для аналізу великих масивів даних, необхідних для професійної діяльності.

**Предмет дисципліни:** методи і технології аналізу великих даних.

**Завдання дисципліни:** формування у студентів системи знань та практичних навичок в галузі використання методів та засобів обробки великих масивів даних; розвиток професійних умінь у сфері інженерії даних та знань щодо розроблення і підтримки комп'ютерних та інформаційних систем, які базуються на інтелектуальному аналізі великих даних.

Після вивчення дисципліни " Аналітика великих даних " студент має досягти таких результатів навчання (сукупність знань, умінь, навичок, компетентностей):

**знати:**

- об'єкт, предмет, задачі, проблематику дисципліни та її основні розділи;
- базові поняття й визначення, використовувані у аналітиці великих даних;
- методологію аналітики великих даних;
- сучасні (в т.ч. й інтелектуальні) методи аналізу великих даних;

**уміти:**

- застосовувати базові концепції і методології моделювання інформаційних процесів у прикладній області;

- виявляти інформаційні потреби і класифікувати дані для проектування інформаційних систем

- оцінювати і застосовувати на системному рівні сучасні програмні та апаратні платформи для ефективного виконання конкретних виробничих задач;

**бути здатним:**

- до абстрактного мислення, аналізу та синтезу; проведення теоретичних та прикладних досліджень на сучасному рівні;

- обирати та використовувати понятійний апарат, методи та інструменти, програмні засоби для здійснення інтелектуального аналізу великих даних (Big Data)

- виконувати самостійний пошук, аналіз та синтез інформації з різних джерел для ефективного рішення спеціалізованих задач професійної діяльності;

- визначати методи інтелектуального аналізу великих даних для розв'язання практичних завдань;

- давати оцінку та інтерпретувати результати аналізу великих даних з точки зору поставленої задачі;

- обирати технології та застосувати їх для виконання інтелектуального аналізу великих даних

- діяти у складних і непередбачуваних умовах, що потребує застосування нових підходів та породження нових ідей (креативності), до саморозвитку та самовдосконалення протягом життя;

- провадити дослідницьку та/або інноваційну діяльність, здійснювати критичний аналіз, оцінювати та синтезувати нові складні ідеї під час аналізу великих даних;

- спілкуватись в діалоговому режимі з широкою науковою спільнотою та громадськістю в галузі аналітики великих даних;

- набувати нові наукові і професійні знання, вдосконалювати навички, прогнозувати розвиток програмних систем та інформаційних технологій;

## СТРУКТУРА ЗАЛКОВИХ КРЕДИТІВ ДИСЦИПЛІНИ

Назва теми	Кількість годин, відведених на:			
	Денна форма			
	Лекції	Практичні роботи	Лабораторні роботи	СРС
<b>Непарний семестр</b>				
Тема 1. Сутність Big Data та інструментарій для обробки та аналізу	6	2	4	18
Тема 2. Основні методи інтелектуального аналізу даних (Data Mining)	22	12	24	111
Тема 3. Основні технології обробки великих даних	6	3	6	26
<b>Разом за семестр:</b>	<b>34</b>	<b>17</b>	<b>34</b>	<b>155</b>

## ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

### Зміст лекційного курсу

№ лекції	Перелік тем лекцій, їх анотації	Кількість годин
	Непарний семестр	
1.	<b>Лекція 1. Big Data і Data Science: сутність та інструментарій.</b> Data Science та Big Data: основні поняття та виклики. Види великих даних. Процес Data Science. Екосистема великих даних та Data Science. Літ.: [1-5]	2
2.	<b>Лекція 2. Класифікація методів аналізу великих даних</b> Основні напрямки роботи з великими даними. Джерела Великих даних. Основні класи методів обробки великих даних. Літ.: [1-5]	2
3.	<b>Лекція 3. Метадані великих даних</b> Метадані різних типів даних. Метадані у соціальних мережах. Метадані сховищ даних. Робота з метаданими. Літ.: [1-5]	2
4.	<b>Лекція 4. Методи очищення даних</b> Збір даних. Підготовка даних. Очищення даних Літ.: [1-5]	2
5.	<b>Лекція 5. OLAP-аналіз</b> Літ.: [1-5]	2
6.	<b>Лекція 6. Методи візуалізації великих даних</b> Дослідження даних. Дашборд: приклади і застосування. Інструменти візуалізації даних Літ.: [1-5]	2
7.	<b>Лекція 7. Методи групування даних</b> Моделювання/аналіз даних. Метрики. Класифікація і кластеризація даних Літ.: [1-5]	2
8.	<b>Лекція 8. Методи групування даних: класифікація даних</b> Класичні технології класифікації. Класифікація методом опорних векторів. Байєсівська класифікація. Класифікація за допомогою методу найближчого сусіда. Класифікація СBR-методом. Алгоритми обмеженого перебору. Класифікація за допомогою штучних нейронних мереж. Класифікація за допомогою дерев рішень. Методи нечіткої логіки. Програмне забезпечення задач класифікації Літ.: [2], с. 250-264; [3], с. 640-751; [4], с. 276-316; [5], [7], [8]	2

9.	<b>Лекція 9. Методи групування даних: кластеризація даних</b> Класичні технології кластеризації. Ієрархічні методи кластерного аналізу. Неієрархічні методи кластерного аналізу. Нові алгоритми кластерного аналізу. Програмне забезпечення задач класифікації Літ.: [2], с. 250-264; [3], с. 640-751; [4], с. 276-316; [5], [7], [8]	2
10.	<b>Лекція 10. Методи аналізу та моделювання залежностей</b> Загальні положення. Статистичні методи. Методи асоціативних правил. Літ.: [1], с. 70-78; [2], с. 180-226; [4], с. 290-301, 370-380, [7]	2
11.	<b>Лекція 11. Секвенційний аналіз</b> Загальна постановка заличАлгоритм AprioriALL. Алгоритм GSP	
12.	<b>Лекція 12 Регресійний аналіз</b> Лінійна регресія. Нелінійна регресія. Множинна регресія	
13.	<b>Лекція 13. Методи аналізу динаміки</b> Основні поняття та визначення. Принципи та етапи аналізу і моделювання динаміки. Програмне забезпечення задач дослідження динаміки. Аналіз та моделювання тенденцій динаміки. Дослідження сезонності Літ.: [1], с. 70-78; [2], с. 180-226; [4], с. 316-370, [7]	2
14.	<b>Лекція 14. Ансамблі моделей</b> Означення. Види ансамблей. Алгоритм Беггінг. Алгоритм бустинга. Алгоритм стекінг. Літ.: [1], с. 70-78; [2], с. 180-226; [4], с. 316-370, [7]	2
15.	<b>Лекція 15. Нейронні мережі та їх застосування при аналізі великих даних</b> Поняття та можливості нейрокомп'ютерних технологій. Нейронні мережі прямого розповсюдження. Метод зворотного розповсюдження помилки. Програмні засоби реалізації нейрокомп'ютерних технологій. Сучасна практика та перспективні напрямки застосування нейротехнологій Літ.: [2], с. 238-250; [3], с. 199-322, [8]	2
16.	<b>Лекція 16. Технології Text Mining</b> Інтелектуальний аналіз тексту. Методи інтелектуального аналізу даних (Data Mining). Методи автоматичного вилучення ключових термінів. Застосування технології Text майнінгу Літ.: [1], с. 257-295; [2], с. 264-301; [3], с. 322-640; [6], [8]	2
17.	<b>Лекція 17 Технології Map Reduce</b> Призначення та області застосування Map Reduce. Основні елементи Map Reduce. Історія розвитку головної технології Big Data. Архітектура Map Reduce. Принцип роботи Map Reduce. Основні переваги моделі Map Reduce. Недоліки і альтернативи Map Reduce. Літ.: [2], с. 315-324; [5], [6], [8]	2
	<b>Разом</b>	34

### Зміст лабораторних занять

№ заняття	Тема лабораторного заняття	Кількість годин
1.	Застосування методів очищення даних.	4
2.	Python: імпорт і попередня обробка даних	4
3.	Python: OLAP-аналіз	4
4.	Класифікація (Дерево рішень)	4
5.	Кластерний аналіз	4
6.	Асоціація	4
7.	Застосування методів аналізу динаміки.	4
8.	Побудова найпростішої нейромережі	6
Разом:		34

### Зміст практичних занять

№ заняття	Тема лабораторного заняття	Кількість годин
1.	Застосування методів очищення даних.	2
2.	Python: імпорт і попередня обробка даних	2
3.	Python: OLAP-аналіз	2
4.	Класифікація (Дерево рішень)	2
5.	Кластерний аналіз	2
6.	Асоціація	2
7.	Застосування методів аналізу динаміки.	2
8.	Штучний інтелект та нейромережі	3*
Разом:		17*

Примітка:

\*-кількість годин практичних занять в залежності від чисельника, знаменника складає 16 або 18

### Зміст самостійної (у т.ч. індивідуальної) роботи

Самостійна робота студентів денної форм навчання полягає у систематичному опрацюванні програмного матеріалу, підготовці до виконання та захисту лабораторних робіт, тестування з теоретичного матеріалу тощо.

Номер тижня	Вид самостійної роботи	К-ть годин
<i>Другий семестр</i>		
1-2	Опрацювання лекційного матеріалу. Підготовка до практичної роботи №1. Підготовка до лабораторної роботи №1	17
3-4	Опрацювання лекційного матеріалу. Підготовка до практичної роботи №2. Підготовка до лабораторної роботи №2	17
5-6	Опрацювання лекційного матеріалу. Підготовка до практичної роботи №3. Підготовка до лабораторної роботи №3	17
7-8	Опрацювання лекційного матеріалу. Підготовка до практичної роботи №4. Підготовка до лабораторної роботи №4.	17
9-10	Опрацювання лекційного матеріалу. Підготовка до практичної роботи №5. Підготовка до лабораторної роботи №5.	17
11-12	Опрацювання лекційного матеріалу. Підготовка до практичної роботи №6. Підготовка до лабораторної роботи №6.	17
13-14	Опрацювання лекційного матеріалу. Підготовка до практичної роботи №7 Підготовка до лабораторної роботи №7.	17
15-16	Опрацювання лекційного матеріалу. Підготовка до практичної роботи №8. Підготовка до лабораторної роботи №8	17
17	Підготовка до контрольної роботи	19

## МЕТОДИ НАВЧАННЯ

Лекції проводяться, в основному, з використанням методів проблемного викладання, словесних, наочних, проблемно-пошукових методів; практичні та лабораторні заняття проводяться пояснювально-ілюстративними методами, методами проблемного викладання, практичними, проблемно-пошуковими, дослідницькими та частково-пошуковими методами; самостійна робота передбачає виконання індивідуальних завдань із залученням методів проблемного викладання, практичних, дослідницьких, частково-пошукових методів.

## ФОРМИ І МЕТОДИ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

Поточний контроль здійснюється під час лекційних та практичних занять, а також у дні проведення контрольних заходів, встановлених робочим планом дисципліни. Семестровий контроль проводиться у формі заліку. При цьому при виведенні остаточної оцінки враховуються результати поточного контролю.

Перед вивченням дисципліни, як правило, проводиться вхідний контроль знань з дисциплін, що їй передують і забезпечують. При цьому необхідно встановити рівні та критерії сформованості знань щодо змісту навчальних елементів. Такими рівнями є:

Ознайомчо-орієнтовний (ОО) – особа має орієнтовне уявлення щодо понять, які вивчаються, здатна: оцінювати якість, надійність, безпеку критичних програмних систем; взаємодіяти та працювати в команді при оцінюванні програмних систем: володіти методами і засобами підтримки командної роботи, планувати та ефективно організовувати роботу, соціальну комунікацію та безперервний контроль якості результатів роботи; діяти у складних і непередбачуваних умовах, що потребує застосування нових підходів, креативності, самостійного пошуку помилок, оцінювання своєї поведінки та постійного самовдосконалення; проводити дослідницьку та/або інноваційну діяльність в галузі оцінювання критичних програмних систем.

Понятійно-аналітичний (ПА) – особа має чітке уявлення щодо навчального об'єкту, здатна перенести раніше засвоєні знання на типові ситуації.

Продуктивно-синтетичний (ПС) – особа має глибоке розуміння щодо навчального об'єкту, здатна здійснювати синтез, генерувати нові ідеї та уявлення, переносити раніше засвоєні знання на нетипові, нестандартні ситуації.

Кожний вид роботи з дисципліни оцінюється за чотирибальною шкалою. Семестрова підсумкова оцінка визначається як середньозважена з усіх видів навчальної роботи, виконаних і зданих позитивно з врахуванням коефіцієнта вагомості. Вагові коефіцієнти змінюються залежно від структури дисципліни і важливості окремих її видів робіт. Студент, який набрав позитивний середньозважений бал за поточну роботу і не здав залік, вважається невстигаючим.

При оцінюванні знань студентів використовуються різні засоби контролю, зокрема: усне опитування перед допуском до виконання практичної та лабораторної роботи – здійснюється на її початку; якість виконання, набуття теоретичних знань і практичних навичок перевіряється шляхом захисту кожної практичної та лабораторної роботи згідно з робочою програмою дисципліни і робочим навчальним планом.

Оцінка, яка виставляється за лабораторне заняття, складається з таких елементів: усне опитування студентів перед допуском до виконання роботи; знання теоретичного матеріалу з теми; якість оформлення звіту; вміння студента обґрунтувати прийняті конструктивні рішення; своєчасний захист роботи.

При оцінюванні знань студентів викладач керується такими критеріями.

Оцінку „зараховано”, за шкалою ECTS – А, отримує студент за глибоке і повне опанування змісту навчального матеріалу, в якому він легко орієнтується, понятійного апарату, за уміння зв'язувати теорію з практикою, вирішувати практичні завдання, висловлювати і обґрунтовувати свої судження. Така оцінка передбачає грамотний, логічний виклад відповіді (як в усній, так і в письмовій формі), якісне зовнішнє оформлення. Студент повинен набути практичних навичок із оцінювання критичних програмних систем.

Оцінка „зараховано”, за шкалою ECTS – А, виставляється студенту, який глибоко засвоїв основні принципи оцінювання критичних програмних систем та вмів раціонально застосовувати їх

при проектуванні програмних систем. Студент не повинен вагатися при видозміні запитання, повинен робити детальні та узагальнюючі висновки.

Оцінку „зараховано”, за шкалою ECTS – В, отримує студент за повне засвоєння навчального матеріалу, володіння понятійним апаратом, орієнтування у вивченому матеріалі, свідоме використання знань для вирішення практичних завдань, грамотний виклад відповіді, але у змісті і формі відповіді мали місце окремі неточності (похибки), нечіткі формулювання закономірностей тощо. Відповідь студента повинна будуватись на основі самостійного мислення.

Оцінку „зараховано”, за шкалою ECTS – С, отримує студент за правильну відповідь з однією-двома суттєвими помилками.

Оцінки „зараховано”, за шкалою ECTS – D, заслуговує студент, який виявив знання основного навчально-програмного матеріалу в обсязі, необхідному для подальшого навчання та практичної діяльності за професією, що справляється з виконанням практичних завдань, передбачених програмою. Як правило, відповідь студента будується на рівні репродуктивного мислення, студент слабо знає структуру курсу, допускає помилки у відповіді, засвоїв і набув практичних навичок з оцінювання критичних програмних систем, але припустився неточностей. Вагається при відповіді на видозмінене запитання, разом з тим студент володіє знаннями, що дозволяють йому під керівництвом викладача усунути неточності у відповіді.

Оцінки „зараховано”, за шкалою ECTS – E, заслуговує студент за неповне опанування програмного матеріалу, але отримані знання і набуті практичні навички із оцінювання якості та експертизи ПЗ.

Оцінка „незараховано”, за шкалою ECTS – FX, виставляється, коли студент має розрізнені, безсистемні знання, не вміє виділяти головне і другорядне, допускається помилок у визначенні понять, перекручує їх зміст, хаотично і невпевнено викладає матеріал, не може використовувати знання при вирішенні практичних завдань. Як правило, оцінка "незараховано", за шкалою ECTS – F, виставляється студенту, який не може продовжити навчання без додаткових знань з курсу.

На основі результатів поточного контролю виставляється підсумкова семестрова оцінка. На основі аналізу контролю знань викладач удосконалює курс лекцій, звертаючи особливу увагу на ті розділи, чи теми, з яких було найбільше неточних відповідей, що свідчить про методичні чи інші недоліки при висвітленні вказаних тем або розділів.

Аналогічно вносяться корективи в методичні посібники для практичних та лабораторних робіт, детальніше розглядаються принципові питання при виконанні практичних та лабораторних робіт та їх захисті.

### Структурування дисципліни за видами робіт і оцінювання результатів навчання студентів денної форми навчання у семестрі за ваговими коефіцієнтами

Аудиторна робота						Самостійна, індивідуальна робота	Залік
<i>II семестр</i>							
Лабораторні роботи №:			Практичні роботи №			Контроль	
1	2	3	1	2	3	Контрольна робота	
ВК:						0,2	
						0,8	

Умовні позначення: Т – тема дисципліни; ВК – ваговий коефіцієнт.

### Структурування дисципліни за видами робіт і оцінювання результатів навчання студентів у семестрі за ваговими коефіцієнтами

2 семестр				
Аудиторна робота				Залік
Лабораторні роботи (6 робіт)	Контрольна робота	Тестовий контроль (3 тести)	Усне опитування	
0,4	0,25	0,25	0,1	-

**Співвідношення вітчизняної шкали оцінювання і шкали оцінювання ЄКТС**

Оцінка ЄКТС	Інституційна шкала балів	Інституційна оцінка	Критерії оцінювання	
A	4,75–5,00	5	Зараховано	<b>Відмінно</b> – глибоке і повне опанування навчального матеріалу і виявлення відповідних умінь та навиків
B	4,25–4,74	4		<b>Добре</b> – повне знання навчального матеріалу з кількома незначними помилками
C	3,75–4,24			<b>Добре</b> – в загальному правильна відповідь з двома-трьома суттєвими помилками
D	3,25–3,74	3		<b>Задовільно</b> – неповне опанування програмного матеріалу, але достатнє для практичної діяльності за професією
E	3,00–3,24	3		<b>Задовільно</b> – неповне опанування програмного матеріалу, що задовольняє мінімальні критерії оцінювання
FX	2,00–2,99	2	Незараховано	<b>Незадовільно</b> – безсистемність одержаних знань і неможливість продовжити навчання без додаткових знань з дисципліни
F	0,00–1,99	2		<b>Незадовільно</b> – необхідна серйозна подальша робота і повторне вивчення дисципліни



## ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ ЗДОБУТИХ СТУДЕНТАМИ ЗНАНЬ

1. Що таке Pandas? Призначення та застосування.
2. Назвіть основні класи методів обробки великих даних. Дайте стислу характеристику.
3. Суть Text Mining та його завдання.
4. Суть Map Reduce та його завдання.
5. Дайте визначення інтелектуального аналізу даних.
6. Що таке розвідувальний аналіз.
7. Що таке якість даних?
8. Які цілі підготовки даних до аналізу?
9. Який атрибут називається цільовим?
10. Що таке значимий і незначний атрибут?
11. Що таке відбір атрибутів?
12. У чому полягає завдання класифікації? Наведіть практичний приклад.
13. Що таке навчання з учителем і без учителя?
14. Завдання класифікації є описовим або прогнозуючим і чому?
15. Навіщо потрібна навчальна і тестова вибірки для вирішення завдання класифікації?
16. Які існують підходи для поділу вихідної вибірки на навчальну і тестову?
17. Метод класифікації 0R.
18. Метод класифікації за одним правилом 1R.
19. Метод класифікації PRISM.
20. Наївна Бассова класифікація NaiveBayes.
21. Метод побудови дерев рішень CART
22. Метод побудови дерев рішень Id3 та C4.5.
23. Метод опорних векторів SMO.
24. Метод k найближчих сусідів kNN.
25. Як оцінити якість побудованої моделі класифікації?
26. Що таке матриця помилок? Як її інтерпретувати?
27. Що означають параметри чутливість, специфічність, точність? Як їх розрахувати?
28. Що таке параметр Каппа? Що він показує?
29. Що таке аналіз витрати-вигоди?
30. Як порівняти роботу двох класифікаторів?
31. Що таке ансамблі класифікаторів і адаптивний бустінг? Для чого вони застосовуються?
32. У чому полягає завдання регресії? Наведіть приклад.
33. Чим завдання регресії схоже і чим відрізняється від завдання класифікації?
34. Завдання регресії є описовим або прогнозуючим і чому?
35. Метод лінійної регресії.
36. Метод опорних векторів для вирішення завдання регресії.
37. Метод найближчих сусідів для вирішення 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
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, які застосовуються в алгоритмі Apriori?
64. Опишіть алгоритм ECLAT
65. Опишіть алгоритм FPGrowth
66. На які дві групи підрозділяються методи Data Mining за принципом роботи з вихідними навчальними даними?
67. Чи правильне таке формулювання: "Асоціація є часткою випадкової послідовності з часовим лагом, що дорівнює нулю"?
68. Що перебуває в основі так званої інформаційної піраміди?
69. На які категорії підрозділяються задачі Data Mining, залежно від моделей, що використовуються?
70. Для вирішення яких задач застосовуються ієрархічні алгоритми?
71. У чому подібність задач класифікації й прогнозування?
72. Який з напрямів включає виявлення закономірностей у діях користувача web-вузла або їхньої групи?
73. Які фактори впливають на результат класифікації в наївнобайєсовському підході?
74. Які переваги характерні для використання байєсовських мереж?
75. Які задачі вирішуються за допомогою методу опорних векторів?
76. Що являє собою група синапсів нейрону?
77. Визначте головну функцію штучного нейрона.
78. У чому полягає навчання мереж, що самоорганізуються?
79. Які існують традиційні методи візуалізації?
80. Назвіть характеристики однієї з основних тенденцій в області візуалізації.
81. Що означає інтегрованість сховища даних?
82. Назвіть основні концепції сховища даних.
83. Якщо набір даних упорядкований та у ньому присутня сезонна або циклічна компонента, то яку мінімальну кількість даних необхідно мати для можливості аналізу?
84. Що повинна робити якісна програма очищення даних?
85. Які характеристики повинна мати якісна програма очищення даних?
86. На якому етапі перетинається робота фахівця предметної області й фахівця з видобутку даних?
87. Назвіть характеристики, властиві SAS Enterprise Miner.
88. Рішення яких задач передбачають алгоритми аналізу даних в PolyAnalyst?
89. Чи існує необхідність тимчасового або постійного копіювання даних для аналізу в системі KXEN?
90. Який компонент KXEN дозволяє виявити природні групи (кластери) у наборі даних?

91. Назвіть слабкі сторони використання готового програмного забезпечення Data Mining.

### РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Черняк О. І. Інтелектуальний аналіз даних : підручник / О. І. Черняк, П. В. Захарченко. – К.: Знання, 2014. – 599 с.
2. Hans Weber. Big Data and Artificial Intelligence: Complete Guide to Data Science, AI, Big Data and Machine Learning. Режим доступу: <https://www.amazon.com/dp/B08FP9YZXN?tag=uuid10-20> .
3. Jay M. Patel. Getting Structured Data from the Internet: Running Web Crawlers/Scrapers on a Big Data Production Scale 1st ed. Edition. Режим доступу: <https://www.amazon.com/dp/1484265750?tag=uuid10-20>.
4. Dursun Delen. Predictive Analytics: Data Mining, Machine Learning and Data Science for Practitioners, 2nd Edition (FT Press Analytics) 2nd Edition. Режим доступу: <https://www.amazon.com/dp/0136738516?tag=uuid10-20>
5. Dinesh Sachdev. Thinking Big: Developers Guide for Big Data Engineering & Analytics. Режим доступу: <https://www.amazon.com/dp/B08TQD9GWQ?tag=uuid10-20>

### ІНФОРМАЦІЙНІ РЕСУРСИ

1. Модульне середовище для навчання MOODLE. Доступ до ресурсу: <https://msn.khnu.km.ua>.
2. Електронна бібліотека університету. Доступ до ресурсу: [http://lib.khnu.km.ua/asp/php\\_f/page\\_lib.php](http://lib.khnu.km.ua/asp/php_f/page_lib.php).