

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ



ЗАТВЕРДЖУЮ

Декан факультету інформаційних технологій  
проф. Тетяна ГОВОРУЩЕНКО

«05» вересня 2024 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Моделювання та інтелектуальна обробка інформації

Назва дисципліни

Галузь знань – 12 Інформаційні технології

Спеціальність – 122 Комп'ютерні науки

Рівень вищої освіти – доктор філософії

Освітньо-наукова програма – Комп'ютерні науки та інформаційні технології

Обсяг дисципліни – 6 кредитів ЄКТС, Шифр дисципліни – ОСП.03

Мова навчання – українська

Статус дисципліни – обов'язкова (спеціальної підготовки)

Факультет - Інформаційних технологій

Кафедра – Комп'ютерних наук

Курс	Семестр	Обсяг дисципліни		Кількість годин						Форма семестрового контролю			
				Аудиторні заняття				Індивідуальна робота студента	Самостійна робота, у т.ч. ІРС	Курсовий проект	Курсова робота	Залік	Іспит
				Разом	Лекції	Лабораторні роботи	Практичні заняття						
1	2	6	180	72	18	36	18		108			+	
<b>Разом</b>		<b>6</b>	<b>180</b>	<b>72</b>	<b>18</b>	<b>36</b>	<b>18</b>		<b>108</b>			<b>1</b>	

Робоча програма складена на основі освітньо-наукової програми «Комп'ютерні науки та інформаційні технології» та стандарту вищої освіти зі спеціальності 122 – Комп'ютерні науки третього (доктор філософії) рівня вищої освіти.

Програма складена \_\_\_\_\_ к.т.н., доц. Руслан БАГРІЙ

Схвалено на засіданні кафедри комп'ютерних наук

Протокол від 30 серпня 2024 р., №1. Зав. Кафедри проф. Олександр БАРМАК

Робоча програма розглянута та схвалена вченою радою факультету інформаційних технологій

Голова вченої ради факультету проф. Тетяна ГОВОРУЩЕНКО

Хмельницький 2024

## МОДЕЛЮВАННЯ ТА ІНТЕЛЕКТУАЛЬНА ОБРОБКА ІНФОРМАЦІЇ

Тип дисципліни	Обов'язкова
Освітній рівень	Третій (доктор філософії)
Мова викладання	українська
Семестр	Другий
Кількість встановлених кредитів ЄКТС	6,0
Форми навчання, для яких викладається дисципліна	денна

**Результати навчання.** Студент, який успішно завершив вивчення дисципліни, повинен: вміло використовувати прийоми структурування вхідної інформації, знати основні підходи до аналізу даних і способи їх модифікації, застосовувати математичні моделі для вирішення завдань аналізу та інтелектуальної обробки інформації, знати множину методів інтелектуального аналізу даних, придатних для обробки інформації в різних її представленнях; вміти аналізувати предметну область, робити узагальнення і синтезувати знання про неї, виокремлювати завдання аналізу даних, визначати їх тип та обирати найбільш відповідний спосіб вирішення, обґрунтовувати оптимальність алгоритмів для інтелектуальної обробки інформації, налаштовувати алгоритми під особливості прикладних задач, вміти працювати з програмними пакетами для аналізу даних та вирішувати конкретні прикладні завдання, чітко і доступно мовою представити результати своїх досліджень у формі доповіді.

**Зміст навчальної дисципліни.** Основні поняття інтелектуальної обробки інформації. Особливості обробки даних. Різноманітні ознаки інформаційних моделей. Задача вибору та генерації інформативної системи ознак. Задача прогнозування. Пошук асоціативних правил. Кластерні технології. Стратегії кластеризації. Багатовимірні евклідові простори. Ієрархічні алгоритми кластеризації. Неієрархічні алгоритми кластеризації. Кластеризація у неевклідових просторах. Адаптивні методи кластеризації. Нечіткі алгоритми кластеризації. Застосування генетичних алгоритмів.

**Пререквізити** – Інтелектуальний аналіз даних, Дискретна математика, Теорія ймовірностей і математична статистика.

**Кореквізити** – кваліфікаційна робота.

**Запланована навчальна діяльність:** лекції – 18 год., лабораторні заняття – 36 год., практичні заняття – 18 год., самостійна робота – 108 год., разом – 180 год.

**Форми (методи) навчання:** лекції (з використанням методів проблемного навчання і візуалізації); лабораторні заняття (з використанням методів комп'ютерного моделювання, тренінгів, майстер-класів, практикумів), самостійна робота (індивідуальні завдання).

**Форми оцінювання результатів навчання:** захист лабораторних робіт; презентація результатів виконання індивідуальних завдань; усне та письмове опитування, підсумковий контрольний захід.

**Вид семестрового контролю:** іспит – 2 семестр.

### Навчальні ресурси:

1. Mohammed J. Zaki, Wagner Meira, Jr., Data Mining and Machine Learning: Fundamental Concepts and Algorithms, 2nd Edition, Cambridge University Press, 2020, 776 pages
2. C. Bouveyron, G. Celeux, T. B. Murphy, and A. E. Raftery, "Preface," in Model-Based Clustering and Classification for Data Science: With Applications in R, Cambridge: Cambridge University Press, 2019, 446 pages.
3. Sánchez, J.S.; García, V. Special Issue on Data Preprocessing in Pattern Recognition: Recent Progress, Trends and Applications. Appl. Sci. 2022
4. Ruano-Ordás, D. Machine Learning-Based Feature Extraction and Selection. Appl. Sci. 2024
5. Benjamin S. Baumer, Daniel T. Kaplan, and Nicholas J. Horton, Modern Data Science with R, 3rd edition, 2024
6. Модульне середовище для навчання MOODLE. Доступ до ресурсу: <https://msn.khnu.km.ua>.
7. Електронна бібліотека університету. Доступ до ресурсу: <http://library.khmnmu.edu.ua/>

**Викладач:** кандидат технічних наук, доцент Багрій Р.О.

## Пояснювальна записка

Дисципліна "Моделювання та інтелектуальна обробка інформації" є дисципліною з циклу спеціальної підготовки докторів філософії в галузі комп'ютерних наук та інформаційних технологій.

Дисципліна «Моделювання та інтелектуальна обробка інформації» має на меті:

- формування у студентів уявлення про типи завдань, що виникають в моделюванні та інтелектуальній обробці інформації.
- вивчення основних підходів та алгоритмів розв'язання задач аналізу та обробки інформації і особливостей їх застосування до вирішення реальних завдань.
- отримання студентами навичок щодо виявлення, формалізації і успішному вирішенню практичних завдань аналізу та обробки інформаційних моделей, що виникають в процесі їх професійної діяльності.
- отримання практичного досвіду в роботі з існуючими програмними пакетами з аналізу та інтелектуальної обробки інформації.

Перед тим як приступити до вивчення дисципліни «Моделювання та інтелектуальна обробка інформації», студенту рекомендується освоїти курси «Інтелектуальний аналіз даних» та «Дискретна математика». Для вивчення розділу, що стосується статистичної теорії розпізнавання образів, студент повинен освоїти курс «Теорія ймовірностей і математична статистика». Крім цього, для успішного освоєння даного курсу студент повинен мати навички самостійної роботи з різними джерелами інформації (Інтернет, друковані видання), умінням узагальнювати інформацію, отриману з різних джерел, умінням представляти результати своїх досліджень.

**Пререквізити** – Інтелектуальний аналіз даних, Дискретна математика, Теорія ймовірностей і математична статистика

**Кореквізити** – кваліфікаційна робота.

Відповідно до Стандарту вищої освіти із зазначеної спеціальності та освітньої програми дисципліна має забезпечити:

### **компетентності:**

ЗК04. Здатність розв'язувати комплексні проблеми комп'ютерних наук на основі системного наукового світогляду та загального культурного кругозору із дотриманням принципів професійної етики та академічної доброчесності.

ФК01. Здатність виконувати оригінальні дослідження, досягати наукових результатів, які створюють нові знання у комп'ютерних науках та дотичних до них міждисциплінарних напрямках і можуть бути опубліковані у провідних наукових виданнях з комп'ютерних наук та суміжних галузей.

ФК02. Здатність застосовувати сучасні методології, методи та інструменти експериментальних і теоретичних досліджень у сфері комп'ютерних наук, сучасні цифрові технології, бази даних та інші електронні ресурси у науковій та освітній діяльності.

ФК03. Здатність виявляти, ставити та вирішувати дослідницькі науково-прикладні задачі та/або проблеми в сфері комп'ютерних наук, оцінювати та забезпечувати якість виконуваних досліджень.

ФК06. Здатність аналізувати та оцінювати сучасний стан і тенденції розвитку комп'ютерних наук та інформаційних технологій.

### **програмні результати навчання:**

ПРН01. Мати передові концептуальні та методологічні знання з комп'ютерних наук і на межі предметних галузей, а також дослідницькі навички, достатні для проведення наукових і прикладних досліджень на рівні останніх світових досягнень з відповідного напрямку, отримання нових знань та/або здійснення інновацій.

ПРН03. Формулювати і перевіряти гіпотези; використовувати для обґрунтування висновків належні докази, зокрема, результати теоретичного аналізу, експериментальних досліджень і математичного та/або комп'ютерного моделювання, наявні літературні дані.

ПРН04. Розробляти та досліджувати концептуальні, математичні і комп'ютерні моделі

процесів і систем, ефективно використовувати їх для отримання нових знань та/або створення інноваційних продуктів у комп'ютерних науках та дотичних міждисциплінарних напрямках.

**ПРН06.** Застосовувати сучасні інструменти і технології пошуку, оброблення та аналізу інформації, зокрема, статистичні методи аналізу даних великого обсягу та/або складної структури, спеціалізовані бази даних та інформаційні системи.

**ПРН08.** Визначати актуальні наукові та практичні проблеми у сфері комп'ютерних наук, глибоко розуміти загальні принципи та методи комп'ютерних наук, а також методологію наукових досліджень, застосувати їх у власних дослідженнях у сфері комп'ютерних наук та у викладацькій практиці.

**Предмет дисципліни.** Методи та алгоритми інтелектуальної обробки інформації в наукових та експериментальних дослідженнях.

**Завдання дисципліни.** Надати студентам знання і практичні навички із застосування методів інтелектуальної аналізу даних, придатних для обробки інформаційних моделей в різних її представленнях.

**Результати навчання.** Студент, який успішно завершив вивчення дисципліни, має досягти таких результатів навчання (сукупність знань, умінь, навичок, компетентностей):

знати:

- прийоми структурування вхідної інформації;
- основні підходи до аналізу даних і способи їх модифікації;
- математичні моделі, що лежать в основі різних підходів до вирішення завдань аналізу та інтелектуальної обробки інформації;
- множину методів інтелектуального аналізу даних, придатних для обробки інформації в різних її представленнях;
- основні підходи для вирішення задач інтелектуального аналізу даних.

вміти:

- аналізувати предметну область, робити узагальнення і синтезувати знання про неї;
- виокремлювати завдання аналізу даних, визначати їх тип та обирати найбільш відповідний спосіб вирішення;
- обґрунтовувати оптимальність алгоритмів для інтелектуальної обробки інформації;
- налаштовувати алгоритми під особливості прикладних задач;
- вміти працювати з програмними пакетами для аналізу даних та вирішувати конкретні прикладні завдання.
- чітко і доступно мовою представити результати своїх досліджень у формі доповіді.

### Структура залікових кредитів дисципліни

Назва теми	Кількість годин, відведених на:			
	лекції	лабор. роботи	прак. роботи	СРС
<b>Другий семестр</b>				
<i>Вступ</i>				
Тема 1. Основні поняття інтелектуальної обробки інформації.	2	4	2	12
<i>Розділ 1. Основна частина</i>				
Тема 2. Особливості обробки даних. Різноманітні ознаки інформаційних моделей.	2	8	2	12
Тема 3. Задача вибору та генерації інформативної системи ознак.	2		2	12
Тема 4. Задача прогнозування.	2	8	2	12
Тема 5. Пошук асоціативних правил.	2		2	12
<i>Розділ 2. Кластерні технології</i>				
Тема 6. Кластерні технології. Стратегії кластеризації. Багатовимірні евклідові простори. Ієрархічні алгоритми кластеризації.	2	8	2	12
Тема 7. Неієрархічні алгоритми кластеризації.	2		2	12
Тема 8. Кластеризація у неевклідових просторах.	2	8	2	12
Тема 9. Адаптивні методи кластеризації. Нечіткі алгоритми кластеризації.	2		2	12
<b>Разом за семестр:</b>	<b>18</b>	<b>36</b>	<b>18</b>	<b>108</b>

### Програма навчальної дисципліни

#### Зміст лекційного курсу\*

Номер лекції	Перелік тем лекцій, їх анотації	Кількість годин
	<i>Другий семестр</i>	
1	<p><b>Вступ. Основні поняття інтелектуальної обробки інформації.</b>                      Задачі обробки даних, як завдання пошуку закономірностей в протоколах спостережень</p> <ul style="list-style-type: none"> <li>– поняття даних, набір даних і їх атрибутів, вимірювання</li> <li>– задачі, етапи та методи аналізу даних</li> <li>– класифікація та властивості методів аналізу даних</li> </ul> <p>Задача класифікації даних.</p> <ul style="list-style-type: none"> <li>– постановка задачі класифікації даних</li> <li>– точність класифікації: оцінка рівня помилок</li> </ul> <p>Задача кластеризації даних.</p> <ul style="list-style-type: none"> <li>– постановка задачі кластеризації даних</li> <li>– статистичний та геометричний підходи</li> <li>– методи оцінки якості рішення і принципи порівняння різних алгоритмів</li> </ul> Літ.: [1, 2, 5]	2
<b>Розділ 1. Основна частина</b>		
2	<p><b>Особливості обробки даних. Різноманітні ознаки інформаційних моделей.</b></p> <ul style="list-style-type: none"> <li>– обчислення відстаней в просторі різноманітних ознак</li> <li>– обчислення відстаней між ознаками в залежності від їх типу</li> <li>– модель MapReduce</li> </ul> Літ.: [1, 2, 3]	2
3	<p><b>Задача вибору та генерації інформативної системи ознак.</b>                      Основні підходи до вирішення задачі обирання інформативної системи</p>	

	<p>доречних ознак (feature selection).</p> <ul style="list-style-type: none"> <li>– критерії якості</li> <li>– алгоритми перебору ознакових підсистем (градієнтні і стохастичні)</li> </ul> <p>Основні підходи до вирішення задачі генерації інформативною системи ознак (feature extraction).</p> <ul style="list-style-type: none"> <li>– метод головних компонент</li> <li>– нелінійні перетворення ознак</li> </ul> <p>Літ.: [1, 2, 4]</p>	
4	<p><b>Задача прогнозування.</b></p> <p>Задача прогнозування (регресійного аналізу), зв'язок геометричної і статистичної постановок.</p> <ul style="list-style-type: none"> <li>– дерева прийняття рішень</li> <li>– побудова лінійної регресії</li> <li>– вплив виду функції втрат (loss function) при відновленні регресійної залежності</li> </ul> <p>Літ.: [1]</p>	2
5	<p><b>Пошук асоціативних правил</b></p> <p>Задача пошуку асоціативних правил</p> <ul style="list-style-type: none"> <li>– асоціативні правила</li> <li>– узагальнені асоціативні правила</li> <li>– чисельні асоціативні правила</li> </ul> <p>Алгоритм Аргіогі та його різновиди</p> <ul style="list-style-type: none"> <li>– властивість анти-монотонності</li> <li>– реалізація алгоритму</li> </ul> <p>Літ.: [1]</p>	2
	<b>Розділ 2. Кластерні технології</b>	
6	<p><b>Кластерні технології. Стратегії кластеризації. Багатовимірні евклідові простори</b></p> <p>Ієрархічні алгоритми кластеризації</p> <ul style="list-style-type: none"> <li>– базовий алгоритм ієрархічної кластеризації</li> <li>– алгоритми Single-link та Complete-link</li> <li>– ефективність ієрархічної кластеризації</li> <li>– альтернативні правила для управління ієрархічною кластеризацією</li> <li>– ієрархічна кластеризація у неевклідових просторах</li> </ul> <p>Літ.: [1, 2, 5]</p>	2
7	<p><b>Неієрархічні алгоритми кластеризації</b></p> <ul style="list-style-type: none"> <li>– алгоритм k-Means (Hard-c-means)</li> <li>– підбір правильного k</li> <li>– алгоритм Бредлі, Файяда та Рейна</li> <li>– переваги представлення N , SUM, SUMSQ</li> <li>– метод найближчого сусіда</li> </ul> <p>Літ.: [1, 2, 5]</p>	2
8	<p><b>Кластеризація у неевклідових просторах</b></p> <ul style="list-style-type: none"> <li>– представлення кластерів у GRGPF-алгоритмі</li> <li>– ініціалізація кластерного дерева</li> <li>– додавання точок у GRGPF-алгоритмі</li> <li>– розділення та об'єднання кластерів</li> <li>– кластеризація у потоках та паралелізм</li> <li>– модель розрахунку потоку</li> <li>– алгоритм кластеризації потоку</li> </ul>	2

	Літ.: [1, 2, 5]	
9	<b>Адаптивні методи кластеризації. Нечіткі алгоритми кластеризації. Застосування генетичних алгоритмів.</b> – алгоритм нечіткої кластеризації – еволюційні методи – застосування генетичних алгоритмів – основні визначення та схема генетичного алгоритму – генетичні операції, представлення даних та відбір Літ.: [1, 2, 5]	2
<b>Разом за семестр:</b>		<b>18</b>

**Примітка.** \*Лекції плануються по 2 год. Якщо у навчальному плані в непарних семестрах запланована 1 год. аудиторних занять на тиждень, то залежно від розкладу занять фактична кількість годин становитиме 36 – по чисельнику, 34 – по знаменнику.

**Зміст лабораторних (практичних, семінарських) занять  
Перелік лабораторних занять для студентів денної форми навчання**

№ з/п	Тема лабораторного заняття	Кількість годин
<i>Другий семестр</i>		
1	Методи первісної обробки даних. Статистичні методи обробки експериментальних даних. Літ.: [6]	4
2	Застосування методів відбору та критеріїв оцінювання інформативних ознак. Літ.: [6]	8
3	Задача прогнозування. Методи побудови асоціативних правил. Літ.: [6]	8
4	Вирішення задачі кластеризації за допомогою ієрархічних та неієрархічних алгоритмів кластеризації. Літ.: [6]	8
5	Вирішення задачі кластеризації за допомогою адаптивних алгоритмів кластеризації. Літ.: [6]	8
<b>Разом за семестр:</b>		<b>36</b>

**Перелік практичних занять для студентів денної форми навчання**

№ з/п	Тема практичного заняття	Кількість годин
<i>Другий семестр</i>		
1	Основні поняття інтелектуальної обробки даних. Літ.: [1, 2, 5]	2
2	Особливості обробки даних. Різноманітні ознаки інформаційних моделей. Літ.: [1, 2, 3]	2
3	Задача вибору та генерації інформативної системи ознак. Літ.: [1, 2, 4]	2
4	Задача прогнозування. Літ.: [1, 2]	2
5	Пошук асоціативних правил Літ.: [1]	2
6	Стратегії кластеризації. Багатовимірні евклідові простори Літ.: [1, 2, 5]	2
7	Неієрархічні алгоритми кластеризації Літ.: [1, 2, 5]	2



8	Кластеризація у неевклідових просторах Літ.: [1, 2, 5]	2
9	Адаптивні методи кластеризації. Нечіткі алгоритми кластеризації. Літ.: [1, 2, 5]	2
<b>Разом за семестр:</b>		<b>18</b>

### Зарахування результатів неформальної освіти

Окремі результати вивчення курсу можуть бути зараховані у випадку отримання студентом результатів навчання у неформальній освіті, що підтверджені відповідним документом (сертифікат, свідоцтво, освітня програма тощо).

Як результат виконання лабораторних робіт №1-2 зараховуються онлайн-курси:

- Курс "Система керування життєвим циклом машинного навчання MLflow"  
<https://app.datacamp.com/learn/courses/introduction-to-mlflow>
- Курс "Інженерія ознак для машинного навчання"  
<https://app.datacamp.com/learn/courses/feature-engineering-with-pyspark>

Як результат виконання лабораторних робіт №3-5 зараховуються онлайн-курси:

- Курс "Машинне навчання з використанням PySpark"  
<https://app.datacamp.com/learn/courses/machine-learning-with-pyspark>

### Зміст самостійної (індивідуальної) роботи

Самостійна робота студентів денної форми навчання полягає у систематичному опрацюванні програмного матеріалу, підготовці до виконання і захисту лабораторних робіт, тестування з теоретичного матеріалу, виконанні індивідуальних завдань тощо.

#### Зміст самостійної роботи студентів денної форми навчання

Тема	Вид самостійної роботи	Кількість годин
1	Загальні поняття про інтелектуальний аналіз даних.	12
2	Детерміновані технології. Імовірнісні технології. Пророблення теоретичного матеріалу та підготовка до лабораторної роботи.	12
3	Бази даних та інтелектуальний аналіз даних. Пророблення теоретичного матеріалу та підготовка до лабораторної роботи.	12
4	Основні моделі технологій інтелектуальних обчислень. Дерева рішень.	12
5	Процес знаходження нового знання. Алгоритми виявлення асоціацій. Пророблення теоретичного матеріалу та підготовка до лабораторної роботи.	12
6	Системи міркування на основі аналогічних випадків. Ієрархічні алгоритми.	12
7	Неієрархічні алгоритми. Пророблення теоретичного матеріалу та підготовка до лабораторної роботи.	12
8	Неевклідова геометрія. Пророблення теоретичного матеріалу.	12
9	Нечітка логіка. Генетичні алгоритми. Пророблення теоретичного матеріалу та підготовка по лабораторної роботи.	12
Всього		108

### Технології навчання

Процес навчання з дисципліни ґрунтується на використанні традиційних та сучасних технологій, зокрема: лекції (з використанням методів візуалізації); лабораторні заняття (з використанням майстер-класів), самостійна робота (індивідуальні завдання).

### Методи контролю

Поточний контроль здійснюється під час лекційних та лабораторних занять, а також у дні проведення контрольних заходів, встановлених робочою програмою і графіком навчального процесу. При цьому використовуються такі методи поточного контролю:



- захист лабораторних робіт;
- презентація і захист індивідуальних завдань.

При виведенні підсумкової семестрової оцінки враховуються результати як поточного контролю, так і підсумкового контрольного заходу, який проводиться у вигляді іспиту, що складається шляхом надання письмових відповідей на питання екзаменаційних білетів. Студент, який набрав позитивний середньозважений бал за поточну роботу і не здав підсумковий контрольний захід (іспит), вважається невестигаючим.

### **Оцінювання результатів навчання студентів у семестрі**

Оцінювання академічних досягнень здобувача вищої освіти здійснюється відповідно до «Положення про контроль і оцінювання результатів навчання здобувачів вищої освіти у ХНУ». Кожний вид роботи з дисципліни оцінюється за інституційною чотирибальною шкалою і виставляється в електронному журналі обліку успішності.

Засвоєння студентом теоретичного матеріалу з дисципліни оцінюється під час захисту лабораторних робіт та підсумковому контрольному заході. Поточний контроль передбачає захист лабораторних робіт. Поточний контроль здійснюється під час лабораторних занять, а також у дні проведення контрольних заходів, встановлених робочою програмою і графіком навчального процесу.

Політика проведення лабораторних робіт з навчальної дисципліни «Моделювання та інтелектуальна обробка інформації»:

1. Студент має бути присутнім на кожному занятті.
2. Студент може бути відсутнім на занятті, якщо здав відповідну лабораторну роботу або з поважної причини: хвороба, тощо.
3. В разі потреби при виконанні лабораторної роботи студент може скористатися засобами зв'язку для пошуку інформації в інтернеті, тощо.
4. На одному занятті захищається тільки одна лабораторна робота. В разі бажання студента здати ще одну роботу, студент має стати в кінець черги. За одне заняття можна здати не більше 2 робіт.
5. Перескладання лабораторної роботи можливе на наступному занятті.
6. Можливе перезарахування деяких лабораторних робіт елементами неформальної освіти згідно Робочої програми дисципліни.

При виконанні лабораторних робіт студент має дотримуватись положень академічної доброчесності; використання напрацювань інших студентів (компоненти програмного забезпечення та програмні коди, графічні матеріали лабораторних робіт та фрагменти звітів з лабораторних робіт) є підставою для зниження оцінки чи перескладання лабораторної роботи.

Студент до виконання лабораторної роботи має бути підготовлений, зокрема: ознайомлений із темою, порядком виконання та відповідними теоретичними відомостями. Для встановлення рівня готовності студента до виконання роботи викладач здійснює опитування.

Пропущене лабораторне заняття студент зобов'язаний відпрацювати не пізніше, ніж за два тижні до кінця теоретичних занять у семестрі.

Семестрова підсумкова оцінка визначається як середньозважена з усіх видів навчальної роботи, виконаних і зданих студентом позитивно, з урахуванням коефіцієнта вагомості і розраховується в автоматизованому режимі за відповідною програмою. Вагові коефіцієнти змінюються залежно від структури дисципліни і важливості окремих видів її робіт. При виведенні підсумкової семестрової оцінки враховуються результати як поточного контролю, так і підсумкового контрольного заходу, який проводиться письмово з усього матеріалу дисципліни. Студент, який набрав позитивний середньозважений бал за поточну роботу і не здав підсумковий контрольний захід, вважається невестигаючим.

Оцінювання знань студентів здійснюється за такими критеріями:

<b>Оцінка за інституційною шкалою</b>	<b>Узагальнений критерій оцінювання</b>
---------------------------------------	---

Відмінно	Здобувач вищої освіти глибоко і у повному обсязі опанував зміст навчального матеріалу, легко в ньому орієнтується і вміло використовує понятійний апарат; уміє пов'язувати теорію з практикою, вирішувати практичні завдання, впевнено висловлювати і обґрунтовувати свої судження. Відмінна оцінка передбачає логічний виклад відповіді державною мовою (в усній або у письмовій формі), демонструє якісне оформлення роботи і володіння спеціальними приладами та інструментами. Здобувач вищої освіти не вагається при видозміні запитання, вміє робити детальні та узагальнюючі висновки. При відповіді допустив дві–три несуттєві похибки.
Добре	Здобувач вищої освіти виявив повне засвоєння навчального матеріалу, володіє понятійним апаратом, орієнтується у вивченому матеріалі; свідомо використовує теоретичні знання для вирішення практичних задач; виклад відповіді грамотний, але у змісті і формі відповіді можуть мати місце окремі неточності, нечіткі формулювання правил, закономірностей тощо. Відповідь здобувача вищої освіти будується на основі самостійного мислення. Здобувач вищої освіти у відповіді допустив дві–три несуттєві помилки.
Задовільно	Здобувач вищої освіти виявив знання основного програмного матеріалу в обсязі, необхідному для подальшого навчання та практичної діяльності за професією, справляється з виконанням практичних завдань, передбачених програмою. Як правило, відповідь здобувача вищої освіти будується на рівні репродуктивного мислення, здобувача вищої освіти має слабкі знання структури дисципліни, допускає неточності і суттєві помилки у відповіді, вагається при відповіді на видозмінене запитання. Разом з тим, набув навичок, необхідних для виконання нескладних практичних завдань, які відповідають мінімальним критеріям оцінювання і володіє знаннями, що дозволяють йому під керівництвом викладача усунути неточності у відповіді.
Незадовільно	Здобувач вищої освіти виявив 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 "незадовільно" виставляється здобувачу вищої освіти, який не може продовжити навчання без додаткової роботи з вивчення дисципліни.

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

Аудиторна робота	Самостійна, індивідуальна робота	Семестровий контроль, іспит
<i>2 семестр</i>		
Лабораторні роботи №:	Індивідуальне завдання:	Підсумковий контрольний захід
ВК: 0,5	0,1	0,4

Умовні позначення: Т – тема дисципліни; ВК – ваговий коефіцієнт, ІЗ – індивідуальне завдання.

Підсумкова семестрова оцінка за інституційною шкалою і шкалою ЄКТС встановлюється в автоматизованому режимі після внесення викладачем усіх оцінок до електронного журналу. Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС наведені у таблиці.

### Співвідношення вітчизняної шкали оцінювання і шкали оцінювання ЄКТС

Оцінка ЄКТС	Інтервальна шкала балів	Вітчизняна оцінка, критерії	
А	4,75–5,00	5	<b>Відмінно</b> – глибоке і повне опанування навчального матеріалу і виявлення відповідних умінь та навиків
В	4,25–4,74	4	<b>Добре</b> – повне знання навчального матеріалу з кількома

			незначними помилками
C	3,75–4,24	4	<i>Добре</i> – в загальному правильна відповідь з двома-трьома суттєвими помилками
D	3,25–3,74	3	<i>Задовільно</i> – неповне опанування програмного матеріалу, але достатнє для практичної діяльності за професією
E	3,00–3,24	3	<i>Задовільно</i> – неповне опанування програмного матеріалу, що задовольняє мінімальні критерії оцінювання
FX	2,00–2,99	2	<i>Незадовільно</i> – безсистемність одержаних знань і неможливість продовжити навчання без додаткових знань з дисципліни
F	0,00–1,99	2	<i>Незадовільно</i> – необхідна серйозна подальша робота і повторне вивчення дисципліни

### Питання для самоконтролю студентів форм навчання

1. Поняття Data Mining.
2. Поняття даних. Набір даних та їх атрибути.
3. Вимірювання даних.
4. Особливості обробки даних.
5. Модель Map Reduce
6. Візуальний аналіз даних. Методи візуалізації
7. Наведіть постановку задачі відбору ознак.
8. Що є результатом виконання процедури відбору ознак?
9. Дайте визначення таких понять: інформативність ознаки, незначущі ознаки, надлишкові ознаки.
10. Проаналізуйте методи відбору інформативних ознак.
11. Порівняйте методи повного перебору.
12. Наведіть переваги та недоліки методів скороченого перебору.
13. У чому полягає основна ідея методу групового врахування аргументів?
14. Які стратегії використовують евристичні методи відбору ознак?
15. Наведіть послідовність виконання методів послідовного додавання та видалення ознак.
16. Які критерії використовуються у методі ранжирування ознак?
17. У яких випадках доцільним є використання методів кластеризації ознак?
18. У чому полягає основна ідея методу випадкового пошуку з адаптацією?
19. Яким чином можуть використовуватися методи еволюційного пошуку до відбору інформативних ознак?
20. Які множини можуть бути використані як початкова точка у методах відбору ознак?
21. Проаналізуйте процедури пошуку оптимального набору ознак.
22. Які стратегії використовуються для оцінювання інформативності набору ознак?
23. Порівняйте критерії оцінювання індивідуальної інформативності.
24. Для оцінювання якого типу зв'язку використовується коефіцієнт парної кореляції?
25. Які критерії використовуються для оцінювання спільного впливу набору ознак?
26. Стандарти Data Mining: CWM
27. Стандарти Data Mining: CRISP
28. Стандарти Data Mining: PMML
29. Задача класифікації
30. Точність класифікації. Оцінка рівня помилок
31. Постановка задачі пошуку асоціативних правил, її різновиди. Представлення результатів.
32. Що таке асоціативне правило? Для чого призначені асоціативні правила?
33. Дати визначення понять підтримки та достовірності правила.
34. Яке призначення алгоритмів пошуку асоціативних правил?
35. На які підзадачі розбивається задача знаходження асоціативних правил?
36. Які методи використовуються для знаходження асоціативних правил?
37. Яким чином обираються значення параметрів  $\text{minsupport}$  та  $\text{minconfidence}$ ?

38. Що таке числові асоціативні правила?
39. Поясніть поняття «узагальнене асоціативне правило».
40. Що називається ієрархією елементів?
41. Які переваги дає введення додаткової інформації прогрупування елементів?
42. Поясніть, які проблеми можуть виникнути при безпосередньому застосуванні алгоритмів знаходження асоціативних правил.
43. В чому полягає сутність виявлення узагальнених асоціативних правил?
44. Яким чином визначають «цікаві» правила? В чому полягає актуальність такого процесу?
45. Дати визначення понять батьківського правила (предка) та найближчого батьківського правила.
46. Порівняйте поняття цікавого та частково цікавого правила.
47. Які проблеми усуває алгоритм обчислення узагальнених асоціативних правил?
48. З яких етапів складається процес обчислення узагальнених асоціативних правил?
49. Проаналізуйте базовий алгоритм пошуку множин, що зустрічаються часто.
50. Опишіть алгоритм генерації кандидатів.
51. Яким чином використовується хеш-дерево для підрахунку підтримки кандидатів? Як відбувається процес побудови такого дерева?
52. Виконайте порівняльний аналіз базового та покращеного алгоритмів пошуку множин, що зустрічаються часто.
53. За рахунок яких оптимізацій відбувається покращення базового алгоритму пошуку множин, що зустрічаються часто?
54. В чому полягає сутність масштабованого алгоритму пошуку асоціативних правил Apriori?
55. Яким чином перетворюються дані для можливості використання алгоритму Apriori?
56. Яка властивість використовується в алгоритмі Apriori? Для чого вона використовується?
57. Наведіть послідовність виконання алгоритму Apriori.
58. Опишіть функцію генерації кандидатів в алгоритмі Apriori.
59. Як відбувається підрахунок підтримки для кожного кандидату в алгоритмі Apriori? Для чого в цій процедурі використовують хеш-дерево?
60. Що таке дерево рішачих правил? Який спосіб подання правил в них використовується?
61. Дати означення основних понять, що відносяться до теорії дерев рішачих правил: об'єкт, атрибут, мітка класу, вузол, лист, перевірка.
62. Навести основні класи задач, до яких можуть бути застосовані дерева рішачих правил.
63. Яким чином відбувається побудова дерева рішачих правил? Який метод використовується для цього?
64. В чому полягає процес навчання з учителем?
65. Порівняйте методи, що реалізують дерева рішачих правил: CART та C4.5.
66. Поясніть принцип роботи «жадібних» алгоритмів.
67. Перелічіть основні аспекти, яким приділяється увага при побудові дерев рішачих правил.
68. В чому полягає правило відбору ознаки для розбиття? Сформулюйте загальне правило для відбору атрибуту.
69. Виконайте порівняльний аналіз критеріїв оцінки якості розбиття множини на класи.
70. Що визначає правило зупину? Дайте порівняльну характеристику відомих критеріїв зупину побудови дерева рішачих правил?
71. Для чого використовується правило відсіку?
72. Що розуміють під точністю та помилкою розпізнавання для дерева рішачих правил?
73. Що необхідно зробити для добування правил з дерева рішачих правил?
74. Постановка задачі кластеризації.
75. Ієрархічні алгоритми кластеризації.
76. Алгоритм k-Means (Hard-c-means).
77. Метод найближчого сусіда.
78. Алгоритм найближчого сусіда.
79. Адаптивні методи кластеризації.

80. Порівняйте методи еволюційного пошуку з іншими методами оптимізації. Які методи відносять до еволюційних?
81. В чому переваги еволюційних методів?
82. Проаналізуйте умови ефективного використання методів еволюційного пошуку.
83. Назвіть особливості еволюційних методів.
84. Які недоліки еволюційного пошуку та в чому вони полягають?
85. Проаналізуйте узагальнену схему роботи еволюційних методів.
86. Наведіть послідовність виконання узагальненого еволюційного пошуку.
87. Які параметри необхідно визначати для роботи еволюційних методів?
88. Які існують способи кодування параметрів, що оптимізуються, при використанні еволюційних методів?
89. Що таке фітнес-функція?
90. Порівняйте стратегії створення початкової популяції.
91. Виконайте порівняльний аналіз операторів відбору (пропорційний відбір, відбір за допомогою ранжирування, турнірний відбір та відбір з використанням порогу).
92. Які способи формування батьківської пари використовуються в еволюційних методах?
93. Проаналізуйте оператори схрещування (точкове, рівномірне, порівняльне, арифметичне).
94. Для чого призначений оператор мутації? Які оператори мутації використовуються в еволюційних методах?
95. Порівняйте класичну із одноточечною мутацією обміну.
96. Яким чином відбувається формування нового покоління?
97. Які критерії зупину використовуються при еволюційному пошуку?

### **Методичне забезпечення**

Навчальний процес з дисципліни «Інтелектуальна обробка інформаційних моделей» повністю і в достатній кількості забезпечений необхідною навчально-методичною літературою.

### **Рекомендована література**

#### **Основна**

1. Mohammed J. Zaki, Wagner Meira, Jr., *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*, 2nd Edition, Cambridge University Press, 2020, 776 pages
2. C. Bouveyron, G. Celeux, T. B. Murphy, and A. E. Raftery, "Preface," in *Model-Based Clustering and Classification for Data Science: With Applications in R*, Cambridge: Cambridge University Press, 2019, 446 pages.
3. Sánchez, J.S.; García, V. Special Issue on Data Preprocessing in Pattern Recognition: Recent Progress, Trends and Applications. *Appl. Sci.* 2022
4. Ruano-Ordás, D. Machine Learning-Based Feature Extraction and Selection. *Appl. Sci.* 2024
5. Benjamin S. Baumer, Daniel T. Kaplan, and Nicholas J. Horton, *Modern Data Science with R*, 3rd edition, 2024
6. Болюбаш Н. М. Інтелектуальний аналіз даних : навч. посіб. / Н. М. Болюбаш. – Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2023. – 320 с.

#### **Допоміжна**

7. Zhang, C.; Li, W.; Zhang, H.; Zhan, T. Recent Advances in Intelligent Data Analysis and Its Applications. *Electronics* 2024
8. Barbierato, E.; Gatti, A. The Challenges of Machine Learning: A Critical Review. *Electronics* 2024.
9. Bhatia, Parteek. *Data mining and data warehousing: principles and practical techniques*. Cambridge, United Kingdom New York, NY: Cambridge University Press, 2019, p. 506.

### **Інформаційні ресурси**

1. Модульне середовище для навчання. Доступ до ресурсу: <https://msn.khnu.km.ua>.
2. Електронна бібліотека університету. Доступ до ресурсу: <http://library.khmnu.edu.ua/>
3. Репозитарій ХНУ. Доступ до ресурсу: <http://elar.khmnu.edu.ua/jspui/?locale=uk>