ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ»
ФАКУЛЬТЕТ МАТЕМАТИКИ ТА ЦИФРОВИХ ТЕХНОЛОГІЙ
Кафедра кібернетики і прикладної математики

«ЗАТВЕРДЖУЮ»
Декан факультету математики
та цифрових технологій
_____________ проф. Микола МАЛЯР
«_30_» ___06___________ 2023 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Інтелектуальні технології Data Mining

Рівень вищої освіти	**перший (бакалаврський)**
Галузь знань	**11 Математика та статистика**
Спеціальність	**113 Прикладна математика**
Освітня програма	**Системи штучного інтелекту**
Статус дисципліни	**обов'язкова**
Мова навчання	**українська**

Ужгород 2023

Робоча програма навчальної дисципліни «**Інтелектуальні технології Data Mining**» для здобувачів вищої освіти галузі знань **11 Математика та статистика** спеціальності **113 Прикладна математика** освітньої програми **Системи штучного інтелекту.**

Розробник: Кондрук Н. Е., доцент, к.т.н, доцент кафедри кібернетики і прикладної математики

Робочу програму розглянуто та затверджено на засіданні кафедри **кібернетики і прикладної математики.**

Протокол № 12 від «05» 06 2023 року.

Завідувач кафедри _____ Павло МУЛЕСА

Схвалено науково-методичною комісією **факультету математики та цифрових технологій.**

Протокол № 10 від «20» червня 2023 року.

Голова науково-методичної комісії _____ Наталія ЮРЧЕНКО

© Кондрук Н.Е., 2023 р.

© ДВНЗ «Ужгородський національний університет», 2023 р.

1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

<table>
<tr><td rowspan="2">Найменування показників</td><td>Розподіл годин за навчальним планом</td></tr>
<tr><td>Денна форма навчання</td></tr>
<tr><td>Кількість кредитів ЄКТС – 5</td><td>Рік підготовки:</td></tr>
<tr><td>Загальна кількість годин – 150</td><td>2-й</td></tr>
<tr><td>Кількість модулів – 2</td><td>Семестр:</td></tr>
<tr><td rowspan="5">Тижневих годин
для денної форми навчання:

аудиторних – 5

самостійної роботи студента – 5</td><td>4-й</td></tr>
<tr><td>Лекції:</td></tr>
<tr><td>32</td></tr>
<tr><td>Практичні (семінарські):</td></tr>
<tr><td></td></tr>
<tr><td rowspan="2">Вид підсумкового контролю: екзамен</td><td>Лабораторні:</td></tr>
<tr><td>42</td></tr>
<tr><td rowspan="2">Форма підсумкового контролю: усна</td><td>Самостійна робота:</td></tr>
<tr><td>76</td></tr>
</table>

2. МЕТА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Метою вивчення навчальної дисципліни «Інтелектуальні технології Data Mining» є вивчення методів інтелектуального аналізу даних, спрямованих на аналітичне дослідження масивів інформації з метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень; огляд методів та алгоритмів, які використовуються в Data Mining; розгляд практичних прикладів застосування Data Mining; підготовка студентів до самостійної роботи з вирішення задач засобами Data Mining.

Відповідно до освітньої програми, вивчення дисципліни сприяє формуванню у здобувачів вищої освіти таких компетентностей:

ЗК01. Здатність учитися і оволодівати сучасними знаннями.
ЗК02. Здатність застосовувати знання у практичних ситуаціях.
ЗК03. Здатність генерувати нові ідеї (креативність).
ЗК04. Здатність бути критичним і самокритичним.
ЗК05. Здатність проведення досліджень на відповідному рівні.
ЗК06. Здатність до абстрактного мислення, аналізу та синтезу.
ЗК08. Знання та розуміння предметної області та розуміння професійної діяльності.
ЗК10. Навички у використанні інформаційних і комунікаційних технологій.
ЗК13. Навички міжособистісної взаємодії.
ЗК16. Здатність до планування та розподілу часу.
ЗК17. Здатність приймати обґрунтовані рішення.
ФК03. Здатність обирати та застосовувати математичні методи для розв’язання прикладних задач, моделювання, аналізу, проектування, керування, прогнозування, прийняття рішень.
ФК04. Здатність розробляти алгоритми та структури даних, програмні засоби та програмну документацію.
ФК06. Здатність розв’язувати професійні задачі за допомогою комп’ютерної техніки, комп’ютерних мереж та Інтернету, в середовищі сучасних операційних систем, з використанням стандартних офісних додатків.
ФК08. Здатність використовувати сучасні технології програмування та тестування програмного забезпечення.
ФК09. Здатність до проведення математичного і комп’ютерного моделювання, аналізу та обробки даних, обчислювального експерименту, розв’язання формалізованих задач за допомогою спеціалізованих програмних засобів.
ФК13. Здатність зрозуміти постановку завдання, сформульовану мовою певної предметної галузі, здійснювати пошук та збір необхідних вихідних даних.
ФК14. Здатність сформулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, та обирати метод її розв’язання, що забезпечує потрібні точність і надійність результату.
ФК17. Здатність адаптовувати методи машинного навчання, інтелектуального аналізу даних для вирішення конкретних задач із різних прикладних областей, проводити відповідні дослідження із аналізом одержаних результатів.

3. ПЕРЕДУМОВИ ДЛЯ ВИВЧЕННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Передумовами вивчення навчальної дисципліни «Інтелектуальні технології Data Mining» є опанування таких навчальних дисциплін (НД) освітньої програми (ОП):

ОК 6. Математичний аналіз
ОК 7. Алгебра і аналітична геометрія
ОК 10. Теорія ймовірностей і математична статистика
ОК 24. Бази даних та інформаційні системи

ОК 15. Вступ до програмування. Python
ОК 20. Операційні системи та стандартні офісні додатки

4. ОЧІКУВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ

Відповідно до освітньої програми **«Системи штучного інтелекту»,** вивчення навчальної дисципліни повинно забезпечити досягнення здобувачами вищої освіти таких програмних результатів навчання (ПРН)**:**

Програмні результати навчання	Шифр ПРН
Демонструвати знання й розуміння основних концепцій, принципів, теорій прикладної математики і використовувати їх на практиці.	РН01
Формалізувати задачі, сформульовані мовою певної предметної галузі; формулювати їх математичну постановку та обирати раціональний метод вирішення; розв'язувати отримані задачі аналітичними та чисельними методами, оцінювати точність та достовірність отриманих результатів	РН03
Вміти проводити практичні дослідження та знаходити розв'язок некоректних задач.	РН07
Володіти методиками вибору раціональних методів та алгоритмів розв'язання математичних задач оптимізації, дослідження операцій, оптимального керування і прийняття рішень, аналізу даних.	РН10
Вміти застосовувати сучасні технології програмування та розроблення програмного забезпечення, програмної реалізації чисельних і символьних алгоритмів.	РН11
Розв'язувати окремі інженерні задачі та/або задачі, що виникають принаймні в одній предметній галузі: в соціології, економіці, екології та медицині.	РН12
Використовувати в практичній роботі спеціалізовані програмні продукти та програмні системи комп'ютерної математики.	РН13
Виявляти здатність до самонавчання та продовження професійного розвитку.	РН14
Демонструвати навички взаємодії з іншими людьми, уміння працювати в команді.	РН16
Вміти обирати, застосовувати та адаптовувати методи Data Mining та машинного навчання для розв'язання прикладних задач.	РН21

Очікувані результати навчання, які повинні бути досягнуті здобувачами освіти після опанування навчальної дисципліни «Інтелектуальні технології Data Mining»:

Очікувані результати навчання з дисципліни	Шифр ПРН
Розуміння основних класів задач інтелектуального аналізу даних.	РН01, РН03 РН14
Застосування відповідних методів інтелектуального аналізу даних до певних класів прикладних задач (за наявності комп'ютерної техніки та відповідного ПЗ), вміння проводити комп'ютерний експеримент реалізований на Python в Google Colaboratory із використанням спеціалізованих бібліотек.	РН01, РН03, РН07, РН10, РН11, РН12, РН13, РН21
Аналіз та оцінка отриманих результатів прикладних досліджень.	РН03, РН07
Застосовувати можливості групової роботи в Гугл Таблицях для розв'язання окремих класів задач інтелектуального аналізу даних.	РН13, РН16,РН21

5. ЗАСОБИ ДІАГНОСТИКИ ТА КРИТЕРІЇ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

Методи навчання

Пояснювально-ілюстративний метод, дослідницький метод, метод кейсів.

Засоби оцінювання та методи демонстрування результатів навчання

Засобами оцінювання та методами демонстрування результатів навчання з навчальної дисципліни є: презентації результатів виконаних лабораторних завдань та їх захист; виконання та захист командних лабораторних робіт; завдання на комп'ютерному обладнанні.

Форми контролю та критерії оцінювання результатів навчання

Форми поточного контролю: виконання та захист лабораторних робіт.
Форма модульного контролю: модульна контрольна робота у формі тестування.
Форма підсумкового семестрового контролю: екзамен.

Розподіл балів, які отримують здобувачі вищої освіти (модуль 1)

Поточне оцінювання та самостійна робота		**Модульна контрольна робота у формі тестування**	**Сума**
Т1	Т2	30	**100**
20	50		

Розподіл балів, які отримують здобувачі вищої освіти (модуль 2)

Поточне оцінювання та самостійна робота			**Модульна контрольна робота у формі тестування**	**Сума**
Т1	Т2	Т3	30	**100**
30	20	20		

Оцінювання окремих видів навчальної роботи з дисципліни

Вид діяльності здобувача вищої освіти	**Модуль 1**		**Модуль 2**	
	Кількість	Максимальна кількість балів (сумарна)	Кількість	Максимальна кількість балів (сумарна)
Лабораторні заняття	4	70	5	70
Модульна контрольна робота		30		30
Разом		**100**		**100**

Критерії оцінювання модульної контрольної роботи

Методика оцінювання. Матеріал кожного модуля, який здобувачі вищої освіти повинні засвоїти протягом семестру, виноситься на одну з двох модульних контрольних робіт, що проводяться в кожному семестрі.

Модульна контрольна робота складається із 10-ти теоретичних тестів, кожне з яких оцінюється в 3 бали.

Невиконані та незахищені лабораторні роботи, а також неявка на модульну контрольну роботу оцінюються в 0 балів незалежно від причини невиконання (неявки).

Сумарна оцінка (від 0 до 100 балів) виставляється у відомість модульного контролю. Модуль зараховується, якщо сумарний бал складає не менше 60 балів, і зараховано всі лабораторні роботи, які є складовими даного модуля.

Здобувач вищої освіти, який не з'явився на модульну контрольну роботу, або ж його модульна оцінка складає від 0 до 34 балів, зобов'язаний скласти (перескласти) модуль до початку підсумкового контролю у строки, визначені викладачем дисципліни та погоджені деканатом факультету.

Критерії оцінювання підсумкового семестрового контролю

Екзаменаційна методика оцінювання. До складання екзамену допускаються здобувачі вищої освіти, які мають підсумковий доекзаменаційний рейтинговий бал не менше 35 і, яким зараховано всі лабораторні роботи.

Підсумкова рейтингова доекзаменаційна оцінка визначається як як середньоарифметичне значення двох модулів.

Здобувач вищої освіти, доекзаменаційний рейтинговий бал якого складає від 0 до 34 балів, зобов'язаний покращити його до початку екзамену під час чергування викладачів на кафедрі у строки, визначені викладачем дисципліни та погоджені деканатом факультету. В протилежному випадку, здобувач не допускається до екзамену, і у нього виникає академічна заборгованість.

Екзамен з навчальної дисципліни здобувач вищої освіти може не складати, якщо він успішно пройшов усі модульні контролі та його влаштовує підсумкова доекзаменаційна рейтингова оцінка за навчальний рік. Здобувачі вищої освіти, рейтинговий бал яких становить від 35 до 59, екзамен складають обов'язково. Здобувач освіти може підвищити на екзамені рейтинговий бал, при цьому, за результатами складання екзамену оцінка не може бути менша за доекзаменаційний рейтинговий бал.

Екзамен проводиться в усній формі. Екзаменаційний білет складається з двох теоретичних питань та одного практичного завдання. Оцінювання результатів навчання на екзамені здійснюється за 100-бальною шкалою. Оцінка за екзамен вноситься у відомість обліку успішності.

<table>
<tr><td rowspan="2">Сума балів за всі види навчальної діяльності</td><td rowspan="2">Оцінка ECTS</td><td colspan="2">Оцінка за національною шкалою</td></tr>
<tr><td>для екзамену, курсового проекту (роботи), практики</td><td>для заліку</td></tr>
<tr><td>90 – 100</td><td>A</td><td>відмінно</td><td rowspan="5">Зараховано</td></tr>
<tr><td>82-89</td><td>B</td><td rowspan="2">добре</td></tr>
<tr><td>74-81</td><td>C</td></tr>
<tr><td>64-73</td><td>D</td><td rowspan="2">задовільно</td></tr>
<tr><td>60-63</td><td>E</td></tr>
<tr><td>35-59</td><td>FX</td><td>незадовільно з можливістю повторного складання</td><td>не зараховано з можливістю повторного складання</td></tr>
<tr><td>0-34</td><td>F</td><td>незадовільно з обов'язковим повторним вивченням дисципліни</td><td>не зараховано з обов'язковим повторним вивченням дисципліни</td></tr>
</table>

Визнання результатів навчання, отриманих у неформальній освіті

Визнання результатів навчання, отриманих у неформальній освіті здійснюється на добровільній основі та передбачає підтвердження того, що здобувач досяг необхідних результатів навчання наявним іменним сертифікатами успішного проходження безоплатного онлайн-курсу «Cluster Analysis in Data Mining» (https://www.coursera.org/programs/program-dierzhavnii-vishchii-navchal-nii-zaklad-uzhghorods-kii-natsional-nii/learn/cluster-analysis?source=search) освітньої платформи Coursera. Таким чином, отриманий сертифікат може бути перезарахований та оцінений на «35» балів. Також отриманий сертифікат курсу «Pattern Discovery in Data Mining» (https://www.coursera.org/learn/data-patterns?specialization=data-mining) може бути перезарахований на «20» балів.

6. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

6.1. Зміст навчальної дисципліни

Модуль 1

Тема 1. Основи технології Data Mining.

Суть, мета та сфера застосування технології Data Mining. Класи систем інтелектуального аналізу даних. Класи задач Data Mining. Поняття датасету. Види атрибутів векторів ознак. Хмарне середовище Google Colaboratory. Очистка та попередня обробка даних. Фільтрація даних. Відновлення пропусків даних. Feature engineering: технології відбору ознак, кодування категоріальних ознак. Перенавчання та недонавчання. Кросвалідація. Бібліотека Scikit learn, Python.

Тема 2. Класифікація.

Моделі задач класифікації. Методи нормалізації даних. Види відстаней між спостереженнями. Метод k найближчих сусідів (kNN). Види метрик точності. Налаштування гіперпараметрів методу для конкретних прикладних задач. Датасет «Іриси Фішера». Наївний Баєсовський класифікатор. Задачі класифікації спаму та розпізнавання рукописних цифр. Дерева прийняття рішень (Decision trees). Ентропія, інформаційний приріст. Алгоритм Ітеративний дихотомайзер (ID3). Індекс Джині. Дерево регресії CART. Налаштування гіперпараметрів методів. Графічна побудова та інтерпретація дерев рішень в Python.

Модуль 2

Тема 1. Кластеризація.

Моделі задач кластеризації. Проблема визначення оптимальної кількості кластерів. Центроїдні методи кластерного аналізу (k-means). Метод ліктя визначення оптимальної кількості кластерів. Ієрархічна кластеризація: агломеративні та дивинізимні методи. Види відстаней між кластерами. Дендрограма. Налаштування гіперпарметрів. Поняття про міри подібності. Змістовна інтерпретація різних видів кластерів.

Тема 2. Прогнозування часових рядів.

Поняття про часові ряди. Методи прогнозування часових рядів. Критерії ефективності прогнозних моделей (MAD, MSE, MAPE, MPE, BIC). Просте ковзаюче середнє (SMA). Експоненціальне згладжування (EMA). Зважене ковзаюче середнє (WMA). Визначення тренду. Кореляційний аналіз. Пряма лінія регресії.

Тема 3. Асоціація.

Асоціативні правила. Метод Apriori. Достовірність (Confidence) та підтримка (Support). Послідовна асоціація, секвенціальний аналіз (SPM). Метод GSP.

6.2. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин											
	Денна форма						Заочна форма					
	Усього	у тому числі					Усього	у тому числі				
		л	п	лаб	інд	ср		л	п	лаб	інд	ср
1	2	3	4	5	6	7	8	9	10	11	12	13
2-й семестр												
Модуль 1.												
Тема 1. Основи технології Data Mining.	16	6		6		4						
Тема 2. Класифікація.	40	8		12		20						
Модульна контрольна робота у формі тестування	8	2				6						
Разом за модуль 1	*64*	*16*		*18*		*30*						
Модуль 2.												
Тема 1. Кластеризація.	32	6		10		16						
Тема 2. Прогнозування часових рядів.	22	4		8		10						
Тема 3. Асоціація.	20	4		6		10						
Модульна контрольна робота у формі тестування	12	2				10						
Разом за модуль 2	*86*	*16*		*24*		*46*						
Разом за семестр	**150**	**32**		**42**		**76**						

6.3. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1	Робота в Google Colaboratory. Попередня обробка даних.	6	
2	Метод найближчих сусідів kNN.	4	
3	Наївний Баєсовський класифікатор	4	
4	Дерева прийняття рішень. ID3, CART.	4	
5	Метод k-середніх. Метод ліктя. Інтерпретація результатів.	4	
6	Ієрархічна кластеризація. Побудова дендрограми. Інтерпретація результатів.	3	
7	Методи прогнозування часових рядів.	8	
8	Метод Apriori. Робота в малих творчих групах.	3	
9	Метод GSP. Робота в малих творчих групах.	3	
	Разом	42	

6.4. Самостійна робота

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1	Основи технології Data Mining. Хмарне середовище Google Colaboratory. Очистка та попередня обробка даних. Feature engineering: технології відбору ознак, кодування ознак. Бібліотека Scikit learn, Python.	4	
2	Класифікація. Метод k найближчих сусідів (kNN). Налаштування гіперпараметрів методу для конкретних прикладних задач. Наївний Баєсовський класифікатор. Дерева прийняття рішень (Decision trees). Налаштування гіперпараметрів методів. Графічна побудова дерев рішень в Python.	20	
3	Підготовка до модульної контрольної роботи №1	6	
4	Кластеризація. Проблема визначення оптимальної кількості кластерів. Центроїдні методи кластерного аналізу (k-means). Метод ліктя Ієрархічна кластеризація. Побудова дендрограма. Просторова кластеризація на основі щільності DBScan. Налаштування гіперпарметрів. Інтерпретація різних видів кластерів.	16	
5	Прогнозування часових рядів. Критерії ефективності прогнозних моделей. Просте ковзаюче середнє (SMA). Експоненціальне згладжування (EMA). Зважене ковзаюче середнє. Кореляційний аналіз. Пряма лінія регресії.	10	
6	Метод Apriori. Визначення асоціативних правил. Метод GSP.	10	
7	Підготовка до модульної контрольної роботи №2	10	
	Разом	76	

7. ІНСТРУМЕНТИ, ОБЛАДНАННЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ, ВИКОРИСТАННЯ ЯКИХ ПЕРЕДБАЧАЄ НАВЧАЛЬНА ДИСЦИПЛІНА

Технічні засоби – персональні комп'ютери, мультимедійний проєктор.
Програмне забезпечення: Гугл Таблиці, Хмарне середовище Google Colaboratory (https://colab.research.google.com/).

1. РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

Основна література

1. Попередня обробка та аналіз даних: лабораторний практикум для студ. спеціальності 113 «Прикладна математика» /Уклад.: Н. Е. Кондрук. Ужгород: УжНУ, 2023. 41 с.
2. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. Харків: ХНУРЕ, 2021. 92 с.
3. Литвин В.В. , Нікольський Ю.В., Пасічник В.В. Аналіз даних та знань. Навчальний посібник. Магнолія, 2021. 276 с.
4. Марченко О. О., Россада Т.В. Актуальні проблеми Data Mining. Київ. 2017. 150 с.

5. Слабоспицький. О. С. Задачі класифікації : навч. посіб. / О. С. Слабоспицький. К. : Видавництво «Людмила», 2020. 43 с.

6. Інтелектуальний аналіз даних. Частина 1 / М.В. Талах, В.В. Дворжак – Чернівці: Технодрук, 2022. 367 с.

7. Лупан І. В. Інтелектуальний аналіз даних Data Mining : навчально-методичний посібник. Кропивницький : ФОП Піскова М. А., 2022. 112 с.

8. Інтелектуальний аналіз даних : практикум / М. Т. Фісун, І. О. Кравець, П. П. Казмірчук, С. Г. Ніколенко. Л. : «Новий світ2000», 2016. 162 с.

9. Kondruk, N.E., Malyar, M.M. Analysis of Cluster Structures by Different Similarity Measures. Cybernetics and Systems Analysis, 2021. 57. Pp. 436–441. URL: https://doi.org/10.1007/s10559-021-00368-4

10. Кондрук Н. Е. Використання мір подібності в методах класифікації. // Науковий вісник Ужгородського університету : серія: Математика і інформатика – Ужгород : Видавництво УжНУ "Говерла", 2021. Вип. 38, №1. с. 143-148. DOI: https://doi.org/10.24144/2616-7700.2021.38(1).143-148

11. Data Mining and Data Warehousing: Principles and Practical Techniques. Front Cover · Parteek Bhatia. Cambridge University Press, 2019. 513 p.

Допоміжна література

1. Ситник В. Ф. Інтелектуальний аналіз даних (дейтамайнінг): Навч. Посібник/ В. Ф. Ситник, М.Т. Краснюк - К: КНЕУ, 2007. - 376 с.

2. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. К.: ІСЗЗІ КПІ ім. Ігоря Сікорського», 2018. 300 с.

3. Черняк О.І.. Інтелектуальний аналіз даних: Підручник / О.І. Черняк, П.В. Захарченко ; Київ. нац. ун-т ім. Т. Шевченка. К. : Знання, 2014. 599 с.

4. North M. Data Mining for the Masses / Matthew North. – Global Text Project, 2012. 264 p.

5. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): навч. посіб. Київ : КНЕУ, 2007. 376 с.

6. Гладун А. Я., Рагушина Ю. В. Data Mining: пошук знань в даних. К.: ТОВ «ВД «АДЕФ Україна», 2016. 452 с.

Інформаційні ресурси в мережі Інтернет

1. Хмарне середовище Google Colaboratory (https://colab.research.google.com/).
2. Візуалізація кластеризації. Доступно з електронного джерела: https://www.naftaliharris.com/blog/visualizing-k-means-clustering/
3. Stanford University Data Mining Lecture Notes– 2020. Режим доступу до ресурсу: http://infolab.stanford.edu/~ullman/mining/2003.html.