

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



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Прикладні системи Data Mining

Рівень вищої освіти	магістерський
Галузь знань	11 Математика та статистика
Спеціальність	113 Прикладна математика
Освітня програма	Науки про дані та інтелектуальні рішення
Статус дисципліни	обов'язкова
Мова навчання	українська

Ужгород 2023

Робоча програма навчальної дисципліни «**Прикладні системи Data Mining**» для здобувачів вищої освіти галузі знань **11 Математика та статистика** спеціальності **113 Прикладна математика** освітньої програми **Науки про дані та інтелектуальні рішення**.

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 «5» червня 2023 р.

Завідувач кафедри  Павло МУЛЕСА.

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

протокол № 10 від «20» червня 2023 р.

Голова науково-методичної комісії  Наталія ЮРЧЕНКО.

© Кондрук Н.Е., 2023 р.

© ДВНЗ «Ужгородський національний університет», 2023 р.

1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Розподіл годин за навчальним планом	
	Денна форма навчання	Заочна форма навчання
Кількість кредитів ЄКТС – 5	Рік підготовки:	
Загальна кількість годин – 150	1-й	-й
Кількість модулів – 2	Семестр:	
Тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи студента – 6	2-й	-й
	Лекції:	
	26	
	Практичні (семінарські):	
Вид підсумкового контролю: екзамен	Лабораторні:	
	34	
Форма підсумкового контролю: усна	Самостійна робота:	
	90	

2. МЕТА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Метою вивчення дисципліни «Прикладні системи майнінгу даних» є формування у студентів компетентностей пов'язаних із теоретичними та прикладними аспектами технологій Data Mining, методами інтелектуального аналізу даних та можливостями їх застосування в умовах розвитку сучасних інформаційних технологій, що сприятиме високій конкурентоспроможності на ринках праці та є основою їх висококваліфікованої професійної діяльності.

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

ЗК01. Здатність до абстрактного мислення, аналізу і синтезу.

ЗК02. Здатність до самонавчання, пошуку, оброблення та аналізу інформації з різних джерел.

ЗК03. Здатність до адаптації та дії в новій ситуації.

ЗК07. Здатність вести професійну, у тому числі науководослідну діяльність у міжнародному середовищі.

ФК02. Здатність обирати та застосовувати математичні методи для розв'язання прикладних задач, моделювання, прогнозування, прийняття рішень, аналізу даних.

ФК03. Здатність до проведення математичного і комп'ютерного моделювання, аналізу та обробки даних, обчислювального експерименту, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів.

ФК04. Здатність розробляти нові та адаптовувати вже існуючі методи та алгоритми розв'язання прикладних задач моделювання та аналізу даних, проводити відповідні експерименти з аналізом одержаних результатів.

ФК05. Здатність розробляти програмне забезпечення для розв'язання формалізованих задач.

ФК07. Здатність пропонувати практичні рішення за фахом з урахуванням сучасних досягнень науки.

ФК08. Здатність будувати нечіткі моделі процесів та явищ у різних галузях науки, розв'язувати задачі обчислювального інтелекту шляхом проектування систем нечіткого виведення, побудови штучних нейронних мереж, розробляти та застосовувати на практиці алгоритми інтелектуального аналізу даних.

3. ПЕРЕДУМОВИ ДЛЯ ВИВЧЕННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Передумовами вивчення навчальної дисципліни «Прикладні системи Data Mining» є базові поняття теорії ймовірностей та математичної статистики, математичного аналізу та опанування таких навчальних дисциплін (НД) освітньої програми (ОП):

ОК 3. Аналітика великих даних

ОК 5. Моделювання нейронних мереж

4. ОЧІКУВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ

Відповідно до освітньої програми «Науки про дані та інтелектуальні рішення», вивчення навчальної дисципліни повинно забезпечити досягнення здобувачами вищої освіти таких програмних результатів навчання (ПРН):

Програмні результати навчання	Шифр ПРН
Використовувати й адаптувати математичні теорії та моделі для забезпечення теоретичного підґрунтя розв'язання наукових та практичних задач.	РН01
Обирати, застосовувати та розробляти нові методи й алгоритми аналізу даних	РН02

для розв'язання наукових та прикладних задач.	
Проводити математичне і комп'ютерне моделювання, аналіз та обробку даних, обчислювальний експеримент, розв'язання формалізованих задач із використанням сучасних методів аналізу даних.	PH04
Формулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, та обирати метод її розв'язання, що забезпечує потрібні точність і надійність результату.	PH05
Уміти будувати комп'ютерний експеримент для конкретних задач прикладної математики шляхом використання спеціалізованих (у тому числі й створених) програмних засобів, та виконувати опис та аналіз результатів експерименту.	PH06

Очікувані результати навчання, які повинні бути досягнуті здобувачами освіти після опанування навчальної дисципліни «Прикладні системи майнінгу даних»:

Очікувані результати навчання з дисципліни	Шифр ПРН
Розуміння основних класів задач інтелектуального аналізу даних, вміння підбирати можливі техніки перетворення вхідних даних для покращення ефективності алгоритмів	PH01, PH02, PH04, PH05
Застосування відповідних методів Data Mining до певних класів прикладних задач (за наявності комп'ютерної техніки)	PH02, PH04, PH05, PH06
Аналіз, оцінка, інтерпретація отриманих результатів прикладних досліджень	PH04, PH06
Навики роботи в малих групах, розподіляти обов'язки та завдання, ефективно планувати свій час, представляти результати досліджень.	PH04, PH06, PH05, PH02,

5. ЗАСОБИ ДІАГНОСТИКИ ТА КРИТЕРІЇ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

Засоби оцінювання та методи демонстрування результатів навчання

Засобами оцінювання та методами демонстрування результатів навчання з навчальної дисципліни є:

командні проекти; реферат, розрахункові та розрахунково-графічні роботи; завдання на комп'ютерному обладнанні, презентації результатів виконаних завдань та досліджень.

Форми контролю та критерії оцінювання результатів навчання

Форми поточного контролю: виконання та захист лабораторних робіт та командних проектів, написання реферату;

Форма модульного контролю: модульне контрольне тестування;

Форма підсумкового семестрового контролю: екзамен.

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

Поточне оцінювання та самостійна робота			Модульне контрольне тестування	Сума
T1	T2	T3	20	100
10	30	40		

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

Поточне оцінювання та самостійна робота		Модульна контрольна робота	Сума
T4	T5	20	100
60	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 (виконання та захист)	4	80	3	60
Реферат	0	0	1	20
Модульна контрольна робота	1	20	1	20
Разом		100		100

Критерії оцінювання модульного контрольного тестування

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.

Модульне тестування складається із 10 тестів, кожне з яких оцінюється в 2 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 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 контролю 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.

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

6. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

6.1. Зміст навчальної дисципліни

Модуль 1

Тема 1. Основи майнінгу даних. Попередня обробка даних.

Основи майнінгу даних. Визначення Data Mining і області застосування. Задачі, моделі та методи Data Mining. Загальні закономірності при аналізі даних. Групи задач аналізу даних. Класифікація методів аналізу. Хмарне середовище Google Colaboratory. Створення CSV датасетів. Очищення та попередня обробка даних. Фільтрація даних. Відновлення пропусків даних. Feature engineering: технології відбору ознак, кодування ознак.

Тема 2. Способи візуального представлення багатовимірних даних та редукції простору ознак. Методи візуалізації. Представлення даних в одному, двох та трьох вимірах в Python. Візуалізація багатовимірних даних. Метод головних компонент (PCA). Пояснювальна дисперсія. Способи визначення оптимальної кількості головних компонент. Нелінійні методи редукції. Т-розподілене вкладення стохастичної близькості (t-SNE). Перплексія. Ізометричне перетворення IsoMAP. Поняття геодезичної відстані. Налаштування гіперпараметрів моделі.

Тема 3. Регресія.

Поняття регресії. Основні етапи регресійного аналізу. Прогнозування неперервної мітки класу. Багатофакторний лінійний аналіз (MLR). Математична модель багатофакторної регресії. Коефіцієнт детермінації. Скоригований коефіцієнт детермінації. Метод Random Forest для задачі регресії. Налаштування гіперпараметрів.

Модуль 2

Тема 4. Моделі класифікації.

Контрольована та неконтрольована класифікація. Математична постановка задач чіткої та нечіткої кластеризації та представлення результатів. Види кластерів. Метрики відстані. Метод Fuzzy c-means. Індеси визначення оптимальної кількості нечітких кластерів. Міри подібності, засновані на нечітких бінарних відношеннях. Методи нечіткого кластерного аналізу засновані на нечітких бінарних відношеннях. Використання підходу структурування даних різними формами кластерів та їх інтерпретація в прикладних задачах. Метричні методи класифікації.

Тема 5. Кібернетичні методи Data Mining.

Еволюційне програмування. Асоціативна пам'ять (пошук аналогів, прототипів). Нечітка логіка; системи обробки експертних знань. Штучні нейронні мережі для задач розпізнавання, класифікації, прогнозу. Генетичні алгоритми.

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 майнінгу даних. Попередня обробка даних.		4		4		4						
Тема 2. Способи візуального представлення даних.		4		8		18						
Тема 3. Регресія.		6		8		18						
Модульне контрольне тестування		1				4						
<i>Разом за модуль 1</i>		<i>15</i>		<i>20</i>		<i>44</i>						
Модуль 2.												
Тема 4. Моделі класифікації.		8		14		20						
Тема 5. Кібернетичні методи Data Mining.		2				22						
Модульне контрольне тестування		1				4						
<i>Разом за модуль 2</i>		<i>11</i>		<i>14</i>		<i>46</i>						
Разом за семестр		26		34		90						

6.3. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1	Особливості роботи в Google Colaboratory. Створення CSV датасетів. Попередня обробка даних.	4	
2	Методи візуалізації. Метод головних компонент (PCA). Т-розподілене вкладення стохастичної близькості (t-SNE). Метричні перетворення IsoMAP.	8	
3	Метод багатофакторної лінійної регресії (MLR).	4	
4	Метод Random Forest.	4	
5	Метод Fuzzy c-means.	4	
6	Робота в малих творчих групах. Методи нечіткого кластерного аналізу засновані на нечітких бінарних відношеннях.	6	
7	Метричні методи класифікації.	4	
	Разом	34	

6.4. Самостійна робота

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1	Основи майнінгу даних. Попередня обробка даних. Хмарне середовище Google Colaboratory. Очистка та попередня обробка даних. Фільтрація даних. Відновлення пропусків даних. Feature engineering: технології відбору ознак, кодування ознак.	4	
2	Способи візуального представлення даних. Представлення даних в одному, двох та трьох вимірах. Візуалізація багатовимірних даних. Метод головних компонент (PCA). Т-розподілене вкладення стохастичної близькості (t-SNE). Самоорганізаційна карта Кохонена.	18	
3	Регресія. Багатофакторний лінійний аналіз (MLR). Метод Random Forest для задачі регресії.	18	
4	Підготовка до модульного контрольного тестування	4	
5	Моделі класифікації. Контрольована та неконтрольована класифікація. Види кластерів. Метрики відстані. Метод Fuzzy c-means. Індeksi визначення оптимальної кількості нечітких кластерів. Міри подібності, засновані на нечітких бінарних відношеннях. Методи нечіткого кластерного аналізу засновані на нечітких бінарних відношеннях. Використання підходу структурування даних різними формами кластерів та їх інтерпретація в прикладних задачах. Метричні методи класифікації.	20	
6	Кібернетичні методи Data Mining. Еволюційне програмування. Асоціативна пам'ять (пошук аналогів, прототипів). Нечітка логіка; системи обробки експертних знань. Штучні нейронні мережі для задач розпізнавання, класифікації, прогнозу. Генетичні алгоритми.	22	

7	Підготовка до модульного контрольного тестування	4	
	Разом	90	

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

Технічні засоби – комп'ютер.

Програмне забезпечення: Colaboratory (<https://colab.research.google.com/>), обране студентом програмне середовище.

8. РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

Основна література

1. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. — Київ. — 2017. — 150 с. URL: <https://cutt.ly/H2S9OD4>
2. Попередня обробка та аналіз даних: лабораторний практикум для студ. спеціальності 113 «Прикладна математика» /Уклад.: Н. Е. Кондрук. Ужгород: УжНУ, 2023. - 41 с.
3. Гладун А. Я., Рагушина Ю. В. Data Mining: пошук знань в даних. К.: ТОВ «ВД «АДЕФ Україна», 2016. 452 с.
4. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с. URL: <https://cutt.ly/j2Dxul5>
5. Слабоспицький. О. С. Задачі класифікації : навч. посіб. / О. С. Слабоспицький. – К. : Видавництво «Людмила», 2020. – 43 с. URL: <https://cutt.ly/12DxCRg>
6. Інтелектуальний аналіз даних : практикум / М. Т. Фісун, І. О. Кравець, П. П. Казмірчук, С. Г. Ніколенко. – Л. : «Новий світ2000», 2016. – 162 с. URL:<https://cutt.ly/q2DBuWw>
7. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. – 92 с. URL: <https://cutt.ly/U2DbE5v>
8. Олійник А. О. Інтелектуальний аналіз даних : навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 278 с. URL: <https://cutt.ly/R2Dmnsd>
9. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних: Підручник, К., 2014. 599 с. URL: <https://cutt.ly/H2DWuan>
10. Data Mining: The Textbook. Front CoverCharu C. Aggarwal. Springer International Publishing, Apr 27, 2015 - Computers - 734 p. URL: <https://cutt.ly/T2DThNX>
11. Data Mining and Data Warehousing: Principles and Practical Techniques. Front Cover · Parateek Bhatia. Cambridge University Press, 2019. – 513 p. URL: <https://cutt.ly/s2DP9Ds>
12. John D. Kelleher (ed.) Fundamentals of Machine Learning for Predictive Data Analytics. Algorithms, Worked Examples, and Case Studies. Cambridge, Mass. : 2020. - URL: <https://cutt.ly/X2DPsvs>
13. Kondruk N. Clustering method based on fuzzy binary relation / N. Kondruk //Eastern-European Journal of Enterprise Technologies. – 2017. – № 2(4). – С. 10–16. DOI: 10.15587/1729–4061.2017.94961
14. Кондрук, Н. Е. Використання довжинної міри подібності в задачах кластеризації /Н. Е. Кондрук //Радіоелектроніка, інформатика, управління. – 2018. – № 3 (46) – С. 98-105. DOI: 10.15588/1607-3274-2018-3-11.
15. Кондрук Н. Е. Аналіз технік зменшення розмірності в машинному навчанні // Науковий вісник Ужгородського університету : серія: Математика і інформатика – Ужгород : Видавництво УжНУ "Говерла", 2023. – Вип. 1 (42). – с. 149 – 155.

16. Kondruk, N.E., Malyar, M.M. Analysis of Cluster Structures by Different Similarity Measures. *Cybernetics and Systems Analysis*, 2021. – 57. Pp. 436–441. <https://doi.org/10.1007/s10559-021-00368-4>
17. 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](https://doi.org/10.24144/2616-7700.2021.38(1).143-148)
18. Кондрук Н. Е. Моделі багатофакторного прогнозування // Науковий вісник Ужгородського університету : серія: Математика і інформатика – Ужгород : Видавництво УжНУ "Говерла", 2022. – Вип. 40, №1 . – с. 168 – 174. DOI [https://doi.org/10.24144/2616-7700.2022.40\(1\).168-174](https://doi.org/10.24144/2616-7700.2022.40(1).168-174)
19. Kondruk, N. E. (2023). Methods for determining similarity of categorical ordered data. *Radio Electronics, Computer Science, Control*, (2), 31. Pp. 31-36. DOI:<https://doi.org/10.15588/1607-3274-2023-2-4> (Web of science)
20. Кондрук, Н. Е. (2023). Аналіз технік зменшення розмірності в машинному навчанні. *Науковий вісник Ужгородського університету. Серія «Математика і інформатика»*, 42(1). С. 181–187. [https://doi.org/10.24144/2616-7700.2023.42\(1\).181-187](https://doi.org/10.24144/2616-7700.2023.42(1).181-187)

Допоміжна література

1. Інтелектуальний аналіз даних: Комп'ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергеев-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського. – Київ : КПІ ім. Ігоря Сікорського, 2018. – 73 с. URL: <https://cutt.ly/M2Dc6b9>
2. Субботін, С. О. Нейронні мережі : навчальний посібник / С. О. Субботін, А. О. Олійник ; під заг. ред. проф. С. О. Субботіна. – Запоріжжя : ЗНТУ, 2014. – 132 с.
3. 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. URL: <http://dwl.kiev.ua/art/oiad/oiad.pdf>
4. Kantarjic M. *Data Mining. Concepts, Models, Methods and Algorithms* / M. Kantarjic, 3rd Ed. – Publisher : Wiley, 2019. – 672 p.
5. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань : навчальний посібник. Львів: «Магнолія 2006», 2015. 276 с.

Інформаційні ресурси в мережі Інтернет

1. Хмарне середовище Google Colaboratory. URL: <https://colab.research.google.com/>
2. Візуалізація кластеризації. URL: <https://www.naftaliharris.com/blog/visualizing-k-means-clustering/>
3. *Data Mining: Practical Machine Learning Tools and Techniques* URL: <https://www.cs.waikato.ac.nz/ml/weka/book.html>
4. *A Programmer's Guide to Data Mining*. URL: <http://guidetodatamining.com/>