

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
Кафедра комп'ютерних наук

ЗАТВЕРДЖУЮ

Декан факультету

Ігор БАРАН

« 02 » 09 20 24 року



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

галузь знань 12 Інформаційні технології

рівень вищої освіти перший рівень (бакалавр)

спеціальність 122 Комп'ютерні науки

освітня програма Комп'ютерні науки

спеціалізація _____

вид дисципліни обов'язкова дисципліна циклу професійної підготовки

Тернопіль 2024

Робоча програма з навчальної дисципліни «Інтелектуальний аналіз даних»
для студентів спеціальності 122 – Комп'ютерні науки факультету комп'ютерно-
інформаційних систем і програмної інженерії.

Розробник: к.т.н., доц. каф. КН Козур Г. / Галина.КОЗБУР

Робоча програма розглянута та схвалена кафедрою комп'ютерних наук,
протокол № 1 від 26.08 2024 р.

Завідувач кафедри комп'ютерних наук [підпис] / Ігор БОДНАРЧУК

Робоча програма розглянута та схвалена НМК факультету комп'ютерно-інформаційних систем і програмної інженерії,

протокол № 1 від « 02.09 20 24 р.

Секретар НМК [підпис] / Богдана МЛИНКО

Робоча програма погоджена:

Спеціальність 122 – Комп'ютерні науки

Освітньо-професійна програма першого рівня вищої освіти за спеціальністю 122
«Комп'ютерні науки» галузі знань 12 «Інформаційні технології»

Гарант освітньої програми [підпис] / Леся ДМИТРОЦА

1. Структура навчальної дисципліни

Показник	Всього годин	
	Денна форма	Заочна форма
Кількість кредитів/годин	4 / 120	4 / 120
Аудиторні заняття, год.	48	12
Самостійна робота, год.	72	108
Аудиторні заняття		
• лекції, год.	16	6
• лабораторні заняття, год.	32	6
• практичні заняття, год.	-	-
• семінарські заняття, год.	-	-
Самостійна робота		
опрацювання матеріалу лекцій	8	18
опрацювання окремих розділів програми, які не виносяться на лекції	6	16
виконання лабораторних робіт	16	32
виконання контрольних завдань	-	-
виконання індивідуальних завдань	-	-
виконання курсових проектів (робіт)	30	30
підготовка до тестування	12	12
Підсумковий контроль	Залік, курсова робота	Залік, курсова робота

Частка годин самостійної роботи студента:

денна форма навчання - 60 %; заочна форма навчання - 90 %.

2. Мета та завдання навчальної дисципліни

2.1 Мета дисципліни «Інтелектуальний аналіз даних» полягає у вивченні сучасних методів та засобів обробки даних, пошуку в масивах структурованих даних раніше невідомих, практично корисних знань та закономірностей, необхідних для прийняття рішень.

2.2 Завдання навчальної дисципліни

За результатами вивчення дисципліни студент повинен продемонструвати такі результати навчання:

- 1) вміння проводити пошук, первинну фільтрацію і очистку «сирих» даних;
- 2) вміння виконувати статистичну обробку великих масивів числових даних з допомогою інструментів електронних таблиць;
- 3) розуміння принципів роботи основних методів і моделей аналізу даних;
- 4) ознайомлення із програмними продуктами та інструментами, що використовуються в Data Mining;
- 5) знання методів валідації побудованих моделей аналізу даних;
- 6) вміння тестувати, аналізувати та створювати візуалізації даних.

3. Компетентності та програмні результати навчання

Вивчення навчальної дисципліни передбачає формування та розвиток у студентів загальних та фахових **компетентностей**.

Інтегральна компетентність: Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми у галузі комп'ютерних наук або у процесі навчання, що передбачає застосування теорій та методів інформаційних технологій і характеризується комплексністю та невизначеністю умов.

Загальні компетентності (КЗ):

КЗ 1. Здатність до абстрактного мислення, аналізу та синтезу;

КЗ 2. Здатність застосовувати знання у практичних ситуаціях;

КЗ 3. Здатність до розуміння предметної області та професійної діяльності;

КЗ 4. Здатність вивчити і оволодіти сучасними знаннями;

ЗК 7. Здатність до пошуку, оброблення та узагальнення інформації з різних джерел;

ЗК 11. Здатність приймати обґрунтовані рішення.

Спеціальні (фахові) компетентності (КС):

СК 11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.

Програмні результати навчання (ПР) згідно освітньої програми:

ПР12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій Data Mining, Text Mining, Web Mining.

4. Опис навчальної дисципліни

4.1. Лекційні заняття

№	Тема заняття та короткий зміст	Кількість годин	
		Денна ф.	Заочна ф.
1.	Вступ до ІАД. Етапи, способи, моделі, методи ІАД. Основні задачі. Дані, їх види. Пошуковий аналіз даних (EDA). Статистичні методи обробки числових даних. Види розподілів даних. Коробкові діаграми.	2	0,5
2.	Основи машинного навчання. Задача класифікації. Дерево рішень як класифікатор. Критерії розбиття на класи. Перенавчання алгоритму.	2	1
3.	Метричні методи класифікації. Метрики якості класифікаторів. Ансамблі дерев рішень. Вибір тренувальних та валідаційних даних для навчання.	2	1
4.	Задача кластеризації. Ітеративні та ієрархічні методи кластеризації.	2	1
5.	Види метрик. Техніки валідації моделі кластеризації.	2	1
6.	Задача регресії. Однофакторна та множинна регресії. Метрики якості для регресійної моделі. Мультиколінеарність.	2	0,5
7.	Знаходження асоціативних правил.	2	0,5
8.	Поняття про інші задачі, методи та засоби інтелектуального аналізу даних. Візуалізація даних.	2	0,5
Усього годин		16	6

4.2. Лабораторні заняття

№	Тема заняття	Кількість годин	
		Денна ф.	Заочна ф.
1.	Пошуковий аналіз даних. Дослідження розподілу даних з допомогою електронних таблиць.	8	1
2.	Аналіз даних з допомогою класифікаційного дерева.	8	2
3.	Побудова моделі кластеризації.	8	2
4.	Знаходження рівняння множинної регресії.	8	1
*	Як альтернатива виконання лабораторного практикуму пропонується освоєння курсів з отриманням сертифікатів на платформі Datacamp.		
Усього годин		32	6

4.3. Самостійна робота

№	Найменування робіт	Кількість годин	
		Денна ф.	Заочна ф.
1.	Опрацювання матеріалу лекцій.	8	18
2.	Опрацювання теоретичного матеріалу, що не виноситься на лекції.	14	16
3.	Виконання та захист лабораторних робіт №№1-4.	8	32
4.	Виконання курсової роботи.	30	30
4.	Підготовка до тесту №1.	6	6
5.	Підготовка до тесту №2.	6	6
Усього годин		72	108

4.4 Перелік тем для самостійного опрацювання

№	Тема	Рекомендовані джерела
1.	Генетичні алгоритми.	[6], с.435-444
2.	Система аналітичного інтерактивного оброблення OLAP.	[6], с.485-491
3.	Ансамблеві математичні моделі ІАД.	[4], с.53-58
4.	Статистична обробка часових рядів.	[10], с.20-33
5.	Text mining, Web mining.	самостійний пошук

5. Курсова робота

Курсова робота студента – заключний етап вивчення дисципліни «Інтелектуальний аналіз даних».

Метою курсової роботи є систематизація та закріплення теоретичних знань, отриманих за час навчання, а також придбання і закріплення навичок самостійної роботи. Курсова робота, як правило, ґрунтується на узагальненні виконаних студентом лабораторних робіт або є індивідуальне дослідницьке завдання з дисципліни і готується до захисту в завершальний період семестру.

Завдання курсової роботи:

- 1) аналіз проблеми та/чи методів і алгоритмів інтелектуального аналізу для її вирішення;
- 2) критичний аналіз існуючих методів і засобів розв'язання аналогічної проблеми;
- 3) обґрунтування вибору методів та/чи засобів розв'язання проблеми;
- 4) виконання поставленого індивідуального завдання та оформлення звітної документації.

Кожна курсова робота є індивідуальною, можлива робота в командах з чітким розмежуванням задач та відповідальності і орієнтована на розвиток у студентів певної частини професійних навичок і вміння творчо вирішувати практичні завдання.

Теми курсових робіт, правила та терміни виконання, оформлення та захисту звітів, а також система оцінювання викладаються в «Методичних вказівках», які розміщено в електронному навчальному курсі. Дасться можливість вільного вибору тематики із запропонованих.

Можлива альтернативна траєкторія виконання КР – через сертифікацію (професійну чи навичкову) на освітній платформі Datacamp в рамках неформальної освіти.

Етапи виконання КР:

1. Вибір та затвердження теми курсової роботи.
2. Критичний аналіз нормативно-правової бази, спеціальної літератури з проблем, що розглядаються, пошук додаткових джерел інформації.
3. Складання плану курсової роботи.
4. Узагальнення та аналіз матеріалу, обробка даних, обґрунтування пропозицій.
5. Написання тексту і оформлення курсової роботи.
6. Захист курсової роботи згідно з встановленим графіком.

Захист курсової роботи передбачає:

- стислу доповідь (5 хв.), в якій необхідно коротко висвітлити зміст одержаних результатів дослідження. Зробити акцент на висновках та рекомендаціях. Можливо, щоб доповідь під час захисту супроводжувалась презентацією результатів, підготовленою за допомогою засобів «Microsoft PowerPoint»;
- співбесіду і відповіді на запитання наукового керівника та членів комісії.

Курсова робота та її захист оцінюється відповідно до вимог кредитно-модульної системи.

6. Критерії оцінювання результатів навчання студентів

Форма підсумкового семестрового контролю – залік.

Підсумкова семестрова оцінка заліку складається з суми балів, отриманих студентом при проходженні двох модульних тестів та отриманих балів за 4 лабораторні роботи та проходження 4 курсів на інтерактивній освітній платформі Datacamp.

За виконання та захист 4-х лабораторних робіт студент отримує максимально 40 балів (10*4).

За проходження Тесту №1 студент максимально отримує 8 балів, Тесту №2 – 7 балів.

За проходження курсів на освітній платформі Datacamp (з отриманням сертифікатів) студент отримує 20 балів (5*4).

Форма підсумкового семестрового контролю – залік (зароблені за семестр бали автоматично домножуються на 4/3).

Модуль 1			Модуль 2			Резом з дисципліни	Залік	Семестрова оцінка
Теорія (тестування)		Практична робота	Теорія (тестування)		Практична робота			
6		10	7		30			
№ лекцій	Вид робіт	Бал	№ лекцій	Вид робіт	Бал	75	*1,33	100
Лекція 1	Лаб. роб. №1	15	Лекція 4,5	Лаб. роб. №3	15			
Лекції 2,3	Лаб. роб. №2	15	Лекція 6	Лаб. роб. №4	15			
			Лекція 7					
			Лекція 8					

Оцінювання курсової роботи

Відповідність курсової роботи вимогам ЄСКД	Викладення змісту розділів	Наявність належно оформлених вступу, висновків, переліку посилань і додатків	Презентація та захист	Сума
5	60	10	25	100

7. Інформаційно-методичне забезпечення

1. Електронний курс «Інтелектуальний аналіз даних» в системі електронного навчання Atutor (ID: 4684, лектор – Козбур Г.В.), який містить:

- документацію до курсу (робоча програма, силабус);
- рекомендації до роботи з електронним курсом;
- актуальний потижневий план проходження дисципліни (основна та альтернативна траскторії);
- теоретичний матеріал та слайди лекцій;
- методичні вказівки до виконання лабораторних робіт та індивідуальні завдання;
- терміни захистів лабораторних робіт та систему оцінювання;
- систему тестування, терміни проходження тренувальних та підсумкових тестів;
- усі актуальні оголошення, опитування, рекомендації;
- додаткову літературу з дисципліни.

8. Рекомендована література

Базова

1. Parteek Bhatia. Data Mining and Data Warehousing. Principles and Practical Techniques. Cambridge University Press. 2009.
2. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms.
3. Han, Jiawei. Data mining : concepts and techniques / Jiawei Han, Micheline Kamber, Jian Pei. – 3rd ed. 2012.
4. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних / Я. В. Іванчук, В. І. Місюра, А. А. Яровий, О. Д. Манжидевський – Вінниця : ВНТУ, 2021. – 69 с.

Додаткова

5. Ian H. Witten. Data Mining. // Academic Press, 2000.
6. Ситник В.Ф. Системи підтримки прийняття рішень: Навч. посібник. – К.: КНЕУ, 2004. – 614 с.
7. E. F. Codd, S. B. Codd, C. T. Salley. Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate. - E. F. Codd & Associates, 1993.
8. J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, H. Pirahesh. Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals // Data Mining and Knowledge Discovery . -- 1997. - N 1. - P. 29-53.
9. Rob P. and C. Coronell. Database Systems: Design, Implementation, and Management, Course Technology, 1997.
10. Данил Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки. навчальний посібник. – К.: ІСТІ КПІ ім. Ігоря Сікорського», 2018. – 297 с.

Електронні інформаційні ресурси

1. <https://www.datacamp.com>
2. <https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/36454/120997.pdf?sequence=2&isAllowed=y>
3. <https://ru.coursera.org/specializations/data-mining>
4. <https://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
5. https://academia.dk/BiologiskAntropologi/Epidemiologi/DataMining/Witten_and_Frank_DataMining_Weka_2nd_Ed_2005.pdf
6. <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>

9. Зміни та доповнення до робочої програми навчальної дисципліни

№	Зміст внесених змін (доповнень)	Дата і № протоколу засідання кафедри	Примітки
1			
2			
3			
4			