МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Львівський національний університет природокористування

Факультет механіки, енергетики та інформаційних технологій

Кафедра інформаційних технологій

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

спеціальність 122 «Комп'ютерні науки»
перший (бакалаврський) рівень вищої освіти

Львів 2024 р.

Робоча програма із дисципліни «Інтелектуальний аналіз даних» для здобувачів першого (бакалаврського) рівня вищої освіти ОП «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки»

Розробник: Тригуба А.М., д.т.н., професор

Робочу програму схвалено на засіданні кафедри інформаційних технологій

Протокол № 1 від 12 серпня 2024 року

Завідувач кафедри інформаційних технологій

(Тригуба А.М.)
(прізвище та ініціали)

Робочу програму схвалено на засіданні методичної комісії факультету механіки, енергетики та інформаційних технологій

Протокол № 1 від 29 серпня 2024 року

Голова методичної комісії факультету механіки, енергетики та інформаційних технологій

(Ковалишин С.Й.)
(прізвище та ініціали)

1. Опис навчальної дисципліни

Галузь знань, спеціальність, освітня програма, рівень вищої освіти,
Галузь знань 12 – інформаційні технології
(шифр і назва)
Спеціальності 122 «Комп'ютерні науки»
Освітня програма «Комп'ютерні науки»
(шифр і назва)
Рівень вищої освіти: перший (бакалаврський)

Характеристика навчальної дисципліни:
Обов'язкова
Кількість кредитів 5
Загальна кількість годин – 150
Індивідуальне науково-дослідне завдання курсова робота
(назва
Вид контролю: екзамен
Тижневих аудиторних годин для денної форми навчання – 4
Співвідношення кількості годин аудиторних занять до самостійної і індивідуальної роботи становить (%):
для денної форми навчання – 59,6%
для заочної форми навчання – 13,6%

2. Програма навчальної дисципліни

Розділ 1. Концептуальні основи інтелектуального аналізу даних.
Тема 1. Предмет дисципліни «Інтелектуальний аналіз даних».
1.1. Мета та завдання дисципліни, її місце в освітньому процесі та в ІТ.
1.2. Поняття інтелектуального аналізу даних.
1.3. Методи, стадії та завдання інтелектуального аналізу даних.
1.4. Data Mining як частина ринку інформаційних технологій.
Тема 2. Препроцесінг інформації.
2.1.Ентропія і кількість інформації.
2.2. Нормалізація і стандартизація вихідних значень.
2.3. Алгоритм „вибілювання" входів.
2.4. Виключення неінформативних чинників.
2.5. Аналітико-евристичні алгоритми визначення вагомих інформативних ознак.
Тема 3. Кореляційний і регресійний аналіз.
3.1. Функціональний і статистичний взаємозв'язки.
3.2. Графічний метод аналізу статистичного взаємозв'язку.
3.3. Форма, тіснота й спрямованість статистичного взаємозв'язку.
3.4. Парний лінійний коефіцієнт кореляції Браве-Пірсона.
3.5. Ранговий коефіцієнт кореляції Спірмена.
3.6. Лінійна регресія.
3.7. Визначення коефіцієнтів рівняння лінійної регресії.

3.8. Дослідження регресійних зв’язків.
3.9. Оцінка достовірності існування статистичного взаємозв’язку.
Тема 4. Поняття та методи класифікації.
4.1. Задачі та види класифікації.
4.2. Методи, що застосовуються для розв’язання задач класифікації.
4.3. Засоби класифікації.
Тема 5. Лінійний дискримінантний аналіз. Побудова канонічних та класифікаційних функцій.
5.1. Методи дискримінації даних.
5.2. Моделі дискримінатного аналізу.
5.3. Лінійний дискримінантний аналіз.
5.4. Побудова канонічних та класифікаційних функцій.
5.5. Зв’язок між дискримінантним, дисперсійним і регресійним аналізом.
Тема 6. Кластеризація. Кластерний аналіз. Ієрархічна та секційна кластеризації.
6.1. Задачі кластеризації.
6.2. Методи та застосування кластерного аналізу
6.3. Кластерний аналіз.
6.4. Методи кластерного аналізу.
6.5. Ієрархічний кластерний аналіз.

Розділ 2. Особливості виконання інтелектуального аналізу даних.

Тема 7. Пошук асоціативних правил. Дерева рішень.
7.1. Основні поняття теорії асоціативних правил.
7.2. Програмні засоби пошуку асоціативних правил.
7.3. Практичний аспект застосування технології асоціативних правил.
7.4. Дерева рішень - загальні принципи технології.
7.5. Комп’ютерні системи та напрямки застосування дерев рішень.
Тема 8. Статистична обробка часових рядів і прогнозування.
8.1. Основні елементи часового ряду.
8.2. Випадкові процеси та часові ряди.
8.3. Розкладення (декомпозиція) часового ряду.
8.4. Прогнозування часових рядів з використанням ARIMA-моделей.
8.5. Прогнозування тенденції на основі зглажування часових рядів.
Тема 9. Візуалізація даних.
9.1. Базові поняття, принципи і цілі візуалізації.
9.2. Метафори і критерії змістовної візуалізації.
9.3. Поняття конвеєра візуалізації як композиції трансформацій прикладних даних.
9.4. Характеристики засобів візуалізації даних.
9.5. Сучасні концепції візуалізації даних.
9.6. Структурні компоненти, принципи і методи підготовки даних.
Тема 10. Аналіз текстової інформації Text Mining.
10.1. Задача та етапи аналізу текстів Text Mining. Загальний опис процесу отримання понять з тексту.

10.2. Методи класифікації текстових документів.
10.3. Методи кластеризації текстових документів: ієрархічні, бінарні.
10.4. Представлення текстових документів. Задача анотування текстів.
10.5. Засоби аналізу текстової інформації.
Тема 11. Добування знань із Web – Web Mining.
11.1. Загальні поняття добування знань з Web - Web Mining.
11.2. Категорії Web Mining.
11.3. Методи витягу Web-контента, а також витягу Web-структур.
Тема 12. Засоби аналізу процесів Process Mining.
12.1. Технологія Process Mining.
12.2. Стандарт запису протоколів MXML.
12.3. Задачі Process Mining.
12.4. Методи Process Mining.
12.5. Інструментальний засіб ProM та його архітектура.

3. Структура навчальної дисципліни

Назви тем	Кількість годин											
	денна форма						заочна форма					
	усього	у тому числі					усього	у тому числі				
		л	п	лаб.	інд.	с. р.		л	п	лаб.	інд.	с. р.
1	2	3	4	5	6	7	8	9	10	11	12	13
	Рік підготовки 3 Семестр 6						Рік підготовки 4 Семестр 7					
Розділ 1. Концептуальні основи інтелектуального аналізу даних												
Тема 1.	6	2	2	–	–	2	6	0	1	–	–	5
Тема 2.	7	2	2	–	–	3	7	1	0	–	–	6
Тема 3.	7	2	2	–	–	3	7	0	1	–	–	6
Тема 4.	7	2	2	–	–	3	7	1	1	–	–	5
Тема 5.	7	2	2	–	–	3	7	1	1	–	–	5
Тема 6.	11	4	4	–	–	3	11	1	1	–	–	9
Разом за розділ 1	45	14	14	–	–	17	45	4	5	–	–	36
Розділ 2. Особливості виконання інтелектуального аналізу даних												
Тема 7.	7	2	2	–	–	3	7	–	1	–	–	6
Тема 8.	6	2	2	–	–	2	6	1	–	–	–	5
Тема 9.	7	2	2	–	–	3	7	–	1	–	–	6
Тема 10.	9	4	2	–	–	3	9	1	1	–	–	7
Тема 11.	7	2	2	–	–	3	7	1	1	–	–	5
Тема 12.	9	2	4	–	–	3	9	1	1	–	–	7
Іспит	30	–	–	–	–	30	30	–	–	–	–	30
Разом за розділ 2	75	14	14	0	0	47	75	4	5	–	–	66
Індивідуальні завдання												
КР	30	–	–	–	–	–	30	–	–	–	–	–
Усього годин	150	28	28	–	–	64	150	8	10	–	–	102

4. Теми лабораторних занять

№ з/п	Назва теми	Кількість,год.
1	Апроксимація. Метод найменших квадратів.	2
2	Кореляційно-регресійний аналіз.	4
3	Інтелектуальний аналіз даних за допомогою програмного пакета WEKA.	2
4	Класифікація і кластеризація.	4
5	Метод найближчих сусідів.	2
6	Встановлення та налаштування для аналізу даних Jupyter Notebook.	4
7	Підготовка даних та розвідувальний аналіз даних.	4
8	Вирішення задачі класифікації.	2
9	Кластерний аналіз даних.	2
10	Добування знань із Web – Web Mining.	2

5. Теми винесені на самостійне вивчення

№ з/п	Назва теми
1	Інструменти Data Mining.
2	Методика «box-counting».
3	Множинна нелінійна регресія. Алгоритм Брандона.
4	Метод опорних векторів.
5	Класифікаційна матриці.
6	Побудова hash-дерев.
7	Алгоритм XGBoost.
8	Засоби аналізу текстової інформації.
9	Хмарні сервіси для візуалізації даних
10	Дослідження використання web-ресурсів.
11	Методи Process Mining.
12	Побудов систем аналізу розподілених даних.

6. Індивідуальні завдання

Тема(и) курсових робіт, завдання – «Використання методів _____(назва) для вирішення _____ (назва задачі) в умовах _____ (предметна галузь).

7. Методи навчання

1. Словесні методи *(лекція, пояснення)*

2. Наочні методи *(презентації, навчальні фільми «Інтелектуальний аналіз даних», «Web – Web Mining»«Text Mining» тощо),*

3. Практичні методи*: лаюораторні роботи, реферати.*

8. Методи контролю:

1. Усне опитування: *фронтальне, індивідуальне.*

2. Письмова аудиторна та позааудиторна перевірка: *рішення задач із інтелектуального аналізу даних, контрольні роботи.*

3. Практична перевірка: *виконання практичних робіт, рішення ситуаційних завдань.*

4. Стандартизований контроль*: тести.*

Види контролю: Поточний контроль, проміжна та семестрова атестація.

9. Результати навчання

У результаті засвоєння окремих тем із дисципліни «Інтелектуальний аналіз даних» здобувачі першого (бакалаврського) рівня вищої освіти набувають знання, уміння та компетентності, що відповідають вимогам ОП «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки».

Індекс в матриці ОПП	**Програмні компоненти**
ЗК1	Здатність до абстрактного мислення, аналізу та синтезу.
ЗК2	Здатність застосовувати знання у практичних ситуаціях.
ЗК7	Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
ЗК8	Здатність генерувати нові ідеї (креативність).
СК2	Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.
СК9	Здатність реалізувати багаторівневу обчислювальну модель на основі архітектури клієнт-сервер, включаючи бази даних, знань і сховища даних, виконувати розподілену обробку великих наборів даних на кластерах стандартних серверів для забезпечення обчислювальних потреб користувачів, у тому числі на хмарних сервісах.
СК11	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.
СК18	Здатність аналізувати масиви даних щодо біологічних об'єктів та процесів природокористування із використанням Data mining, створювати штучні нейронні мережі для вирішення інтелектуальних задач регресії, класифікації, кластеризації та асоціації, а також на їх основі обґрунтовувати рішення, виконувати передбачення та здійснювати управління.
ПРН1	Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.
ПРН3	Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.
ПРН4	Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.
ПРН12	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 DataMining, TextMining, WebMining.
ПРН18	Застосовувати знання для розв'язання складних спеціалізованих завдань інтелектуальної комп'ютеризації у сфері природокористування та інтелектуального аналізу даних в процесі професійної діяльності, в тому числі щодо оцінки стану біологічних об'єктів та виконання процесів природокористування на підставі застосування сучасних методів, моделей, алгоритмів машинного навчання та штучних нейронних мереж.

10. Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота (разом 50балів)												Підсумковий контроль (екзамен)	Сума
розділ 1						розділ 2							
Т=1	Т2	Т3	Т4	Т5	Т6	Т7	Т8	Т9	Т10	Т11	Т12	50 балів	100
3	4	4	4	4	4	4	4	4	4	4	4		

Т1, Т2 ... Т12 – теми

11. Методичне забезпечення

Підручник, навчальний посібник; методичні рекомендації до практичних занять; текстові та електронні варіанти тестів для поточного і підсумкового контролю, методичні матеріали для організації самостійної роботи студентів.

12. Рекомендована література

Базова

1. Гладун А. Я., Рагушина Ю. В. Data Mining: пошук знань в даних. К.: ТОВ «ВД «АДЕФ Україна», 2016. 452 с.
2. Ситник В. Ф.,Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посібник. К: КНЕУ, 2007. 376 с.
3. Sarkar D., Bali R., Sharma T. Practical Machine Learning with Python. Apress, 2018. 545 p. URL: www.apress.com/978-1-4842-3206-4.
4. Zaki M. J., Meira W. Jr. Data mining and analysis : Fundamental Concepts and Algorithms. New York : Cambridge University Press, 2014. 604 p.

Допоміжна

1. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань : навчальний посібник. Львів: «Магнолія 2006», 2015. 276 с.
2. Tryhuba, A., Koval, N., Tryhuba, I., Boiarchuk, O. Application of Sarima Models in Information Systems Forecasting Seasonal Volumes of Food Raw Materials of Procurement on the Territory of Communities. CEUR Workshop Proceedings. 2022. 3295, p. 64-75.
3. Tryhuba A., Boyarchuk V., Tryhuba I., Ftoma O., Padyuka R., Rudynets M. Forecasting the risk of the resource demand for dairy farms basing on machine learning (MoMLeT&DS-2020) In: CEUR Workshop Proceedings, 2020, 2631, pp. 327-340.
4. Бахрушин В. Є. Методи аналізу даних : навчальний посібник для . Запоріжжя: КПУ, 2011. 268 с.
5. Інтелектуальний аналіз даних : Комп’ютерний практикум : навч. посібник / О. О. Сергеєв-Горчинський, Г. В. Іщенко. К. : КПІ ім. Ігоря Сікорського, 2018. 73 с.
6. Інтелектуальний аналіз даних : практикум / М. Т. Фісун, І. О. Кравець, П. П. Казмірчук, С. Г. Ніколенко. Л. : «Новий світ-2000», 2016. 162 с.
7. Снитюк В. Є. Прогнозування. Моделі. Методи. Алгоритми : навчальний посібник. К.: Маклаут, 2008. 364 с.
8. Литвин В.В., Пасічник В.В., Яцишин Ю.В. Інтелектуальні системи : підручник. Львів: Новий світ – 2000, 2009. 406с.

9. Математичне забезпечення інформаційно-керуючих систем: підручник / А. М. Гуржій, З. В. Дудар, В. М. Левикін, Б. В. Шамша. Х. : Компанія Сміт, 2006. 448 с.
10. Vikram Dayal. An Introduction to R for Quantitative Economics: Graphing, Simulating and Computing. Springer, 2015. ISBN 978-81-322-2340-5. http://www.springer.com/978-81-322-2340-5
11. K. Soetaert, J. Cash, and F. Mazzia. Solving Differential Equations in R. Use R! Springer, 2012. ISBN 978-3-642-28070-2
12. Kantarzic M. Data Mining. Concepts, Models, Methods and Algorithms / M. Kantarzic, 3rd Ed. Publisher : Wiley, 2019. 672 p.

15. Інформаційні ресурси

1. Бібліотечно-інформаційні ресурси – книжковий фонд, періодика та фонди на електронних носіях бібліотеки ЛНУП, державних органів науково-технічної інформації, наукових, науково-технічних бібліотек та інших наукових бібліотек України.
2. Віртуальне навчальне середовище ЛНУП. URL: https://moodle.lnup.edu.ua/
3. Електронні інформаційні ресурси мережі інтернет:

- Прискорена гібридна хмарна платформа даних. URL: https://www.weka.io/accelerated-cloud-data-platform/
- JupyterLab. URL: https://jupyter.org/
- Evolutionary Complexity Research Group (EPlex): http://eplex.cs.ucf.edu/
- Матеріали відкритого курсу OpenDataScience [Електронний ресурс]. Електрон. дан. Режим доступа: World Wide Web. URL: https://habr.com/ru/company/ods/blog/344044.
- The latest in machine learning. Papers With Code [Електронний ресурс]. Електрон. дан. Режим доступу: World Wide Web. URL: https://paperswithcode.com/.
- Платформа для змагань з аналітики та передбачувального моделювання. [Електронний ресурс]. Режим доступу: https://www.kaggle.com/.
- Портал відкритих даних України. URL: https://data.gov.ua/
- Weka Machine learning software to solve data mining problems [Електронний ресурс]. URL: https://sourceforge.net/projects/weka/?source=typ_redirect.