

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТАВРІЙСЬКИЙ ДЕРЖАВНИЙ АГРОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ДМИТРА МОТОРНОГО**

Кафедра «Комп'ютерні науки»

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри КН

доцент  Юлія ХОЛОДНЯК

“31” серпня _2022 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Data Science та Data mining»

для здобувачів ступеня вищої освіти «Бакалавр»
зі спеціальності 051 «Економіка»
за ОПП Економічна кібернетика та програмування
(на основі повної загальної середньої освіти)

факультет економіки та бізнесу

2022– 2023 н.р.

Робоча програма навчальної дисципліни «Data Science та Data mining» для здобувачів ступеня вищої освіти «Бакалавр» зі спеціальності 051 «Економіка» за ОПП Економічна кібернетика та програмування (на основі повної загальної середньої освіти) - Запоріжжя, ТДАТУ ім. Д. Моторного. - 12 с.

Розробник: д.т.н., професор Малкіна В.М

Робоча програма затверджена на засіданні кафедри «Комп'ютерні науки»

Протокол від № 1 від 31 серпня 2022 року

В.о. завідувача кафедри КН

доцент  Юлія ХОЛОДНЯК

Схвалено методичною комісією факультет економіки та бізнесу зі спеціальності 051 «Економіка» за ОПП Економічна кібернетика та програмування (на основі повної загальної середньої освіти)

Протокол № 1 від «2» вересня 2022 року

Голова, доц.  Анна КОСТЯКОВА

1 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, спеціальність, ступінь вищої освіти	Характеристика навчальної дисципліни	
		денна форма навчання (денна або заочна)	
Кількість кредитів 4	Галузь знань 12 "Інформатика та обчислювальна техніка"	Обов'язкова	
Загальна кількість годин –120 годин	Спеціальність 051 «Економіка»	Курс	Семестр
Змістових модулів – 2		4-й	7-й
Тижневе навантаження: - аудиторних занять 4 год. - самостійна робота студента 8 год.	Ступінь вищої освіти «Бакалавр»	Вид занять	Кількість годин
		Лекції	20 год.
		Лабораторні заняття	-
		Практичні заняття	20 год.
		Семінарські заняття	-
		Самостійна робота	80 год.
		Форма контролю: <u>екзамен</u>	

2 МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета дисципліни “ Data Science та Data mining ” - вивчення методів сучасної обробки даних – інтелектуального аналізу даних (Data Mining), пошуку у необроблених масивах даних раніше невідомих, практично корисних знань та закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів та різних інструментальних засобів, що використовуються Data Mining; розгляд практичних прикладів застосування Data Mining; підготовка студентів до самостійної роботи з розв'язання різних економічних задач засобами Data Mining та розробки інтелектуальних систем. Розглядаються такі загальні поняття: статистичні пакети; нейронні мережі; еволюційні методи і алгоритми пошуку логічних закономірностей.

Завдання курсу - навчитися застосовувати сучасні інформаційні технології і програмні системи для обробки експериментальних і статистичних даних.

Вивчення навчальної дисципліни передбачає формування наступних компетентностей студента:.

Результати навчання (з урахуванням soft skills):.

Інтегральна компетентність

Здатність розв'язувати задачі в галузі професійної та/або дослідницько-інноваційної діяльності у сфері комп'ютерних наук, що передбачає як вільне володіння наявними знаннями, так і спроможність їх застосування у професійній практиці

Загальні компетенції

Здатність до абстрактного мислення, аналізу та синтезу.

Здатність застосовувати знання у практичних ситуаціях.

Знання та розуміння предметної області та розуміння професійної діяльності.

Здатність спілкуватися іноземною мовою.

Здатність вчитися й оволодівати сучасними знаннями.

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

Здатність приймати обґрунтовані рішення.

Здатність генерувати нові ідеї (креативність)

Здатність оцінювати та забезпечувати якість виконуваних робіт.

Здатність діяти на основі етичних міркувань.

Здатність реалізувати свої права і обов'язки як члена

суспільства, усвідомлювати цінності громадянського (вільного демократичного) суспільства та необхідність його сталого розвитку, верховенства права, прав і свобод людини і громадянина в Україні.

Спеціальні (фахові, предметні) компетентності.

Здатність до математичного формулювання та досліджування неперервних та дискретних математичних моделей, обґрунтування вибору методів і підходів для розв'язування теоретичних і прикладних задач у галузі комп'ютерних наук, аналізу та інтерпретування

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.

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.

Soft skills:

- **комунікативні навички:** письмове, вербальне й невербальне спілкування; уміння грамотно спілкуватися по e-mail; вести суперечки і відстоювати свою позицію, спілкування в конфліктній ситуації; навички створення, керування й побудови відносин у команді;

- **уміння виступати привселюдно:** навички, необхідні для виступів на публіці; проводити презентації;

- **керування часом:** уміння справлятися із завданнями вчасно;

- **гнучкість і адаптивність:** гнучкість, адаптивність і здатність мінятися; уміння аналізувати ситуацію, орієнтування на вирішення проблем;

- **лідерські якості:** уміння спокійно працювати в напруженому середовищі; уміння ухвалювати рішення; уміння встановлювати мету, планувати;

- **особисті якості:** креативне й критичне мислення; етичність, чесність, терпіння, повага до колег

3 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1

Тема 1 Технології аналізу даних [1, 6-29; 3, 42-61]

- 1 Вступ до аналізу даних
- 2 Принципи аналізу даних
- 3 Методика отримання знань
- 4 Поняття алгоритмів data mining

Тема 2 Інтелектуальний аналіз даних на основі регресії [1, 23-38; 3, 42-61]

- 1 Регресія, основні поняття
- 2 Проста лінійна регресія

Тема 3 Регресійний аналіз [1, 69-74; 3, 75-95]

- 1 Проста лінійна регресійна модель
- 2 Гіпотези в регресійному аналізі
- 3 Оцінка значимості регресійної моделі: t-критерій і f-критерій
- 4 Множинна лінійна регресія
- 5 Побудова та аналіз моделі множинної лінійної регресії
- 6 Оцінка значимості множинної регресійної моделі
- 7 Обмеження при застосуванні регресійної моделі

Тема 4 Аналіз і прогнозування часових рядів [1, 29-67; 3, 65-85]

- 1 Поняття прогнозування
- 2 Поняття часового ряду
- 3 Моделі часових рядів
- 4 Автокореляція при дослідженні часових рядів
- 5 Моделі прогнозування часових рядів
- 6 Прогнозування на основі методу декомпозиції часового ряду

Змістовий модуль 2

Тема 5 Класифікація. машинне навчання. Деревя рішень [1, 105-109; 2, 70-92]

- 1 Основні поняття дерев рішень
- 2 Алгоритми побудови дерев рішень

Тема 6 Кластеризація [1, 112-125; 2, 110-145]

- 1 Основні поняття

- 2 Методи кластерного аналізу
- 3 Алгоритм кластеризації k-means
- 4 Приклад роботи алгоритму k-means
- 5 Оцінка якості багатовимірної класифікації

Тема 7 Задача асоціації. [1, 169-174; 2, 175-195]

Асоціативні правила. Пошук асоціативних правил. Генерація асоціативних правил. Послідовні шаблони. Алгоритм Аргіогі. Алгоритм FPG.

4. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Номер тижня	Вид занять	Тема заняття або завдання на самостійну роботу	Кількість				
			годин				балів
			лк	лаб	сем. (пр.)	СРС	
Змістовий модуль 1.							
1	Лекція 1	Технології аналізу даних	2	-	-	-	-
	Лабораторна робота 1	Первинна обробка статистичних даних. Описова статистика	-	2	-	-	3
	Самостійна робота 1	Можливості системи Statistica. Перевірка статистичних гіпотез Приклади застосування Data Mining.	-	-	-	8	2
2	Лекція 2	Основні поняття статистичного аналізу даних	2	-	-	-	-
	Лабораторна робота 2	Проста лінійна регресія	-	2	-	-	3
	Самостійна робота 2	Архітектура, класифікація та стандарти програмних систем дейтамайнінгу	-	-	-	8	2
3	Лекція 3	Класифікація і регресія	2	-	-	-	-
	Лабораторна робота 3	Нелінійна регресія	-	2	-	-	3
	Самостійна робота 3	Характеристика сучасного стану ринку систем дейтамайнінгу та напрями розвитку	-	-	-	8	2
4	Лекція 4	Регресійний аналіз	2	-	-	-	-
	Лабораторна робота 4	Множинна регресія	-	2	-	-	3

	Самостійна робота 4	Характеристики процесів активностей методів дейтамайнінгу	-	-	-	8	2
5	Лекція 5	Аналіз прогнозування часових рядів	2	-	-	-	-
	Лабораторна робота 5	Аналіз часових рядів	-	2	-	-	3
	Самостійна робота 6	Розроблення програмного забезпечення штучних нейромереж	-	-	-	8	2
6,7	ПМК 1	Підсумковий контроль за змістовий модуль 1					10
Всього за змістовий модуль 1 - 55 год.			10	10	-	40	35

Змістовий модуль 2.							
8	Лекція 6	Класифікація, машинне навчання, дерево рішень	2	-	-	-	-
	Лабораторна робота 6	Класифікація за допомогою дерева рішень. Алгоритм CART	-	2	-	-	3
	Самостійна робота 6	Програмне забезпечення нейромереж	-	-	-	8	2
9	Лекція 7	Класифікація. Алгоритм покриття	2				
	Лабораторна робота 7	Класифікація. Алгоритм покриття.		2			3
	Самостійна робота 7	Підготовка до лабораторної роботи				8	2
10	Лекція 8	Data Mining: кластеризація	2	-	-	-	-
	Лабораторна робота 8	Дерево рішень. Алгоритм ID3	-	2	-	-	3
	Самостійна робота 7	Програмне забезпечення нейромереж	-	-	-	8	2
11	Лекція 9	Задачі асоціації	2	-	-	-	-

	Лабораторна робота 9	Задача кластеризації	-	2	-	-	3
	Самостійна робота 8	Загальна схема генетичних алгоритмів	-	-	-	8	2
12	Лекція 10	Консолідація даних	2				
	Лабораторна робота 10	Пошук асоціативних правил. Алгоритм Аргіогі		2			3
	Самостійна робота 10	Підготовка до лабораторної роботи				8	2
13,14	ПМК 2	Підсумковий контроль за змістовий модуль 2					10
Всього за змістовий модуль 2 - 65год.			10	10	-	40	35
Всього з навчальної дисципліни - 120 год.							100

5. ПЕРЕЛІК ПИТАНЬ, ЩО ВІНОСЯТЬСЯ НА ПІДСУМКОВИЙ МОДУЛЬНИЙ КОНТРОЛЬ №1

1. Сутність аналітичних технологій.
2. Поняття інтелектуального аналізу даних.
3. Етапи та методи пошуку нових знань.
4. Основні моделі інтелектуальних обчислень.
5. Засоби програмної підтримки інтелектуального аналізу даних.
6. Новітні напрями застосування Data Mining.
7. Концепція сховищ даних.
8. Технології побудови сховищ даних.
9. Вітрини та кіоски даних .
10. OLAP-технологія.
11. Основні архітектури OLAP-систем.
12. OLAP-системи та інтернет-технології.
13. Поняття та можливості нейрокомп'ютерних технологій.
14. Архітектура нейронних мереж.
15. Нейронні мережі Хопфілда та Кохонена.
16. Програмні засоби реалізації нейромережових технологій.
17. Сучасна практика та перспективні напрями застосування нейротехнологій.

ПЕРЕЛІК ПИТАНЬ, ЩО ВІНОСЯТЬСЯ НА ПІДСУМКОВИЙ МОДУЛЬНИЙ КОНТРОЛЬ №2

1. Основні поняття теорії асоціативних правил.
2. Програмні засоби пошуку асоціативних правил.

3. Практичний аспект застосування технології асоціативних правил.
4. Дерева рішень — загальні принципи технології.
5. Комп'ютерні системи та напрямки застосування дерев рішень.
6. Концептуальні засади еволюційної теорії. Основні положення теорії генетичних алгоритмів.
7. Моделі генетичних алгоритмів.
8. Програмне забезпечення та сфери застосування генетичних алгоритмів.
9. Концепція нечітких обчислень.
10. Нечітка логіка в системах Data Mining.
11. Програмне забезпечення нечітких методів.
12. Сучасна практика застосування нечітких методів.

6. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Базова

1. Ситник В. Ф. Інтелектуальний аналіз даних (дейтамайнінг): Навч. Посібник/ В. Ф. Ситник, М.Т. Краснюк - К: КНЕУ, 2007. - 376 с.
2. 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.
3. Ситник В.Ф. Засоби дейтамайнінгу для аналізу бізнесових рішень. Науково-практичний журнал "Науково-технічна інформація", №3, 2002. - с. 60-64.
4. Ситник В.Ф. Системи підтримки прийняття рішень: Навч.посібник / В.Ф. Ситник - К.:КНЕУ, 2004. – 614 с.

Допоміжна

- 1 Єремєєв В.С. Теорія ймовірностей та математична статистика/В.С. Єремєєв, Д.О. Сосновських, О.В. Тітова. - Навчальний посібник: Мелітополь: ТОВ «Видавничий будинок ММД», 2009. – 188 с.
- 2 Han J. Data Mining: Concepts and Techniques (Second Edition)/ J. Han, M. Kamber – Morgan Kaufmann Publishers, 2006. – 800 p.
- 3 Witten, I.H. Data mining : practical machine learning tools and techniques.—3rd ed. / Ian H. Witten, Frank Eibe, Mark A. Hall. – Morgan Kaufmann Publishers, 2011. – 629 p

7. ІНФОРМАЦІЙНІ РЕСУРСИ

- 1 Освітній портал ТДАТУ <http://op.tsatu.edu.ua>
- 2 Освітній портал ТДАТУ
<http://op.tsatu.edu.ua/course/view.php?id=827#section-8>
- 3 Сайт кафедри КН <http://www.tsatu.edu.ua/kn>