

ПОЛТАВСЬКИЙ УНІВЕРСИТЕТ ЕКОНОМІКИ І ТОРГІВЛІ
Навчально-науковий інститут денної освіти
Кафедра комп'ютерних наук та інформаційних технологій

ЗАТВЕРДЖУЮ

Завідувач кафедри КНІТ

 Олена ОЛЬХОВСЬКА

« 28 » __06__ 2024 р.

РОБОЧА ПРОГРАМА

навчальної дисципліни
«Статистичної обробки»

освітня програма

спеціальність

галузь знань

ступінь вищої освіти

«Аналіз даних та прикладні пакети

Комп'ютерні науки

122 Комп'ютерні науки

12 Інформаційні технології

бакалавр

Робоча програма навчальної дисципліни «Аналіз даних та прикладні пакети статистичної обробки» рекомендована до використання в освітньому процесі на засіданні кафедри комп'ютерних наук та інформаційних технологій
Протокол від 28.06.2024 року, №13

Полтава 2024


Укладачі:

Олексійчук Юрій Федорович, доцент кафедри комп'ютерних наук та інформаційних технологій, к. ф.-м. н.

Оріхівська Оксана Григорівна, старший викладач кафедри комп'ютерних наук та інформаційних технологій

ПОГОДЖЕНО:

Гарант освітньої програми «Комп'ютерні науки»
спеціальності 122 "Комп'ютерні науки"
ступеня бакалавр

 Оксана ЧЕРНЕНКО
підпис ініціали, прізвище

«_28_» ____06____ 2024 року

Розділ 1. Опис навчальної дисципліни

Таблиця 1. Опис навчальної дисципліни "Аналіз даних та прикладні пакети статистичної обробки"

Місце у структурно-логічній схемі підготовки	<i>Пререквізити:</i> – Теорія ймовірностей і математична статистика. <i>Постреквізити:</i> –
Мова викладання	Українська
Статус дисципліни	Обов'язкова
Курс/семестр вивчення	4/7
Кількість кредитів ECTS/кількість модулів	4/2
Денна форма навчання:	
Кількість годин: 120 год	
- Лекції: 16 год	
- Практичні (семінарські, лабораторні) заняття: 32 год	
- Самостійна робота: 72 год.	
- Вид підсумкового контролю (ПМК, екзамен): ПМК	
Заочна форма навчання:	
Кількість годин: 120 год	
- Лекції: 8 год	
- Практичні (семінарські, лабораторні) заняття: 4 год	
- Самостійна робота: 108 год	
- Вид підсумкового контролю (ПМК, екзамен): ПМК	

Розділ 2. Перелік компетентностей, які забезпечує дана навчальна дисципліна, програмні результати навчання

Метою вивчення дисципліни «Аналіз даних та прикладні пакети статистичної обробки» є формування уявлення про методи статистичної обробки даних із використанням комп'ютерних технологій для розв'язування практичних задач; виробити навички вибору відповідних статистичних моделей та методів для наявних даних.

Таблиця 2. Перелік компетентностей та програмні результати навчання, які забезпечує навчальна дисципліна «АДППСО»

Програмні результати навчання	Компетентності, якими повинен оволодіти здобувач
ПР1. 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	Здатність до абстрактного мислення, аналізу та синтезу (ЗК1). Здатність застосовувати знання у практичних ситуаціях (ЗК2). Знання та розуміння предметної області та

комп'ютерних наук.

ПР3. 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.

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 DataMining, TextMining, WebMining.

розуміння професійної діяльності (ЗК3).

Здатність спілкуватися державною мовою як усно, так і письмово (ЗК4).

Здатність вчитися й оволодівати сучасними знаннями (ЗК6).

Здатність до пошуку, оброблення та аналізу інформації з різних джерел (ЗК7).

Здатність генерувати нові ідеї (креативність) (ЗК8).

Здатність працювати в команді (ЗК9).

Здатність бути критичним і самокритичним (ЗК10).

Здатність приймати обґрунтовані рішення (ЗК11)

Здатність оцінювати та забезпечувати якість виконуваних робіт (ЗК12).

Здатність діяти на основі етичних міркувань (ЗК13).

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 (СК2).

Здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів систем, проводити обчислювальні експерименти з обробкою й аналізом результатів (СК7).

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 (СК11).

Розділ 3. Програма навчальної дисципліни

Модуль 1. Основи аналізу даних та методи Data Mining

Тема 1. Попередній аналіз

Основні задачі аналізу даних. Підготовка вхідної інформації. Класифікація змінних. Попередня обробка. Аномальні спостереження. Робота з пропущеними значеннями. Розвідувальний аналіз. Гіпотези та їх перевірка. Дисперсійний аналіз. Огляд статистичних пакетів та бібліотек для аналізу даних. Основи візуалізації (гістограми, діаграми розсіювання, коробкові діаграми). Інструменти для візуалізації.

Тема 2. Кореляційний аналіз

Функціональна та стохастична залежності. Коефіцієнт кореляції Пірсона. Частинна кореляція. Множинний коефіцієнт кореляції. Кореляція непараметричних даних. Коефіцієнт кореляції Фехнера. Коефіцієнт рангової кореляції Спірмена. Кореляція Кендала. Візуалізація кореляцій (теплові карти).

Тема 3. Регресійний аналіз

Проста лінійна регресія. Множинна лінійна регресія. Нелінійна регресія. Проблема мультиколінеарності. Перевірка якості моделі. Прогнозування.

Тема 4. Класифікація та кластеризація даних

Задача класифікації. Задача кластеризації. Метрики оцінки. Дискримінантний аналіз. Логістична регресія. Метод k-найближчих сусідів. Відстань і схожість між об'єктами. Методи кластеризації. Кластеризація методом k-середніх (k-means).

Модуль 2. Інтелектуальний аналіз даних з використанням статистичних пакетів

Тема 5. Інтелектуальний аналіз даних

Data Mining. Text Mining. Web Mining. Бібліотека scikit-learn. Векторизація тексту. Зменшення розмірності даних. Методи аналізу веб-даних. Частотний аналіз слів. Аналіз тональності тексту. Рекомендаційні системи.

Тема 6. Пошук асоціативних правил

Основи задачі пошуку асоціативних правил: підтримка, достовірність, приріст. Алгоритм Аргіої. Алгоритм FP-росту (FP-Growth).

Розділ 4. Тематичний план навчальної дисципліни

Таблиця 4. Тематичний план навчальної дисципліни для студентів денної форми навчання

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	кількість годин	Назва теми та питання семінарського, практичного або лабораторного заняття	кількість годин	Завдання самостійної роботи в розрізі тем	кількість годин
Модуль 1. Основи аналізу даних та методи Data Mining					
Тема 1. Попередній аналіз					
Лекція 1-2. Попередній аналіз Основні задачі аналізу даних. Підготовка вхідної інформації. Класифікація змінних. Попередня обробка. Аномальні спостереження. Робота з пропущеними значеннями. Розвідувальний аналіз. Гіпотези та їх перевірка.	4	Практична робота 1. Основні статистичні характеристики, виявлення аномальних спостережень Практична робота 2. Аналіз даних в Python. Візуалізація даних	2 2	опрацювати лекційний матеріал до теми 1, готуватись до практичних занять	12

1	2	3	4	5	6
Дисперсійний аналіз. Огляд статистичних пакетів та бібліотек для аналізу даних. Основи візуалізації (гістограми, діаграми розсіювання, коробкові діаграми). Інструменти для візуалізації.					
<u>Тема 2. Кореляційний аналіз</u>					
Лекція 3. Кореляційний аналіз Функціональна та стохастична залежності. Коефіцієнт кореляції Пірсона. Частинна кореляція. Множинний коефіцієнт кореляції. Кореляція непараметричних даних. Коефіцієнт кореляції Фехнера. Коефіцієнт рангової кореляції Спірмена. Кореляція Кендала. Візуалізація кореляцій (теплові карти).	2	Практична робота 3. Кореляція кількісних та якісних змінних. Практична робота 4. Частинна кореляція. Множинна кореляція.	2 2	опрацювати лекційний матеріал до теми 2, готуватись до практичних занять	12
<u>Тема 3. Регресійний аналіз</u>					
Лекція 4. Регресійний аналіз Проста лінійна регресія. Множинна лінійна регресія. Нелінійна регресія. Проблема мультиколінеарності. Перевірка якості моделі. Прогнозування.	2	Практична робота 5. Лінійна регресія Практична робота 6. Нелінійна регресія Практична робота 7. Мультиколінеарність	2 2 2	опрацювати лекційний матеріал до теми 3, готуватись до практичних занять	12
<u>Тема 4. Класифікація та кластеризація даних</u>					
Лекція 5. Класифікація та кластеризація даних Задача класифікації. Задача кластеризації. Метрики оцінки. Дискримінантний аналіз. Логістична регресія. Метод k-найближчих сусідів.	2	Практична робота 8. Метод k-найближчих сусідів Практична робота 9. Логістична регресія Практична робота 10. Методи кластеризації	2 2 2	опрацювати лекційний матеріал до теми 4, готуватись до практичних занять	12

1	2	3	4	5	6
Відстань і схожість між об'єктами. Методи кластеризації. Кластеризація методом k-середніх (k-means).		Практична робота 11. Контрольна робота №1	2		
Модуль 2. Інтелектуальний аналіз даних з використанням статистичних пакетів					
<u>Тема 5. Інтелектуальний аналіз даних</u>					
Лекція 6-7. Інтелектуальний аналіз даних Data Mining. Text Mining. Web Mining. Бібліотека scikit-learn. Векторизація тексту. Зменшення розмірності даних. Методи аналізу веб-даних. Частотний аналіз слів. Аналіз тональності тексту. Рекомендаційні системи.	4	Практична робота 12. Text Mining Практична робота 13. Web Mining Практична робота 14. Рекомендаційна система	2 2 2	опрацювати лекційний матеріал до теми 5, готуватись до практичних занять	12
<u>Тема 6. Пошук асоціативних правил</u>					
Лекція 8. Пошук асоціативних правил Основи задачі пошуку асоціативних правил: підтримка, достовірність, приріст. Алгоритм Аргіогі. Алгоритм FP-росту (FP-Growth).	2	Практична робота 15. Пошук асоціативних правил Практична робота 16. Контрольна робота №2	2 2	опрацювати лекційний матеріал до теми 6, готуватись до практичних занять	12
Всього, годин	16		32		72

Таблиця 4. Тематичний план навчальної дисципліни для студентів заочної форми навчання

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	кількість годин	Назва теми та питання семінарського, практичного або лабораторного заняття	кількість годин	Завдання самостійної роботи в розрізі тем	кількість годин
Модуль 1. Основи аналізу даних та методи Data Mining					
<u>Тема 1. Попередній аналіз</u>					
Лекція 1-2. Попередній аналіз Основні задачі аналізу	2	Практична робота 1. Основні статистичні характеристики,	2	опрацювати лекційний матеріал	16

1	2	3	4	5	6
даних. Підготовка вхідної інформації. Класифікація змінних. Попередня обробка. Аномальні спостереження. Робота з пропущеними значеннями. Розвідувальний аналіз. Гіпотези та їх перевірка. Дисперсійний аналіз. Огляд статистичних пакетів та бібліотек для аналізу даних. Основи візуалізації (гістограми, діаграми розсіювання, коробкові діаграми). Інструменти для візуалізації.		виявлення аномальних спостережень Практична робота 2. Аналіз даних в Python. Візуалізація даних		до теми 1, готуватись до практичних занять	
<u>Тема 2. Кореляційний аналіз</u>					
Лекція 3. Кореляційний аналіз Функціональна та стохастична залежності. Коефіцієнт кореляції Пірсона. Частинна кореляція. Множинний коефіцієнт кореляції. Кореляція непараметричних даних. Коефіцієнт кореляції Фехнера. Коефіцієнт рангової кореляції Спірмена. Кореляція Кендала. Візуалізація кореляцій (теплові карти).	2	Практична робота 3. Кореляція кількісних та якісних змінних. Практична робота 4. Частинна кореляція. Множинна кореляція.		опрацювати лекційний матеріал до теми 2, готуватись до практичних занять	16
<u>Тема 3. Регресійний аналіз</u>					
Лекція 4. Регресійний аналіз Проста лінійна регресія. Множинна лінійна регресія. Нелінійна регресія. Проблема мультиколінеарності. Перевірка якості моделі. Прогнозування.		Практична робота 5. Лінійна регресія Практична робота 6. Нелінійна регресія Практична робота 7. Мультиколінеарність		опрацювати лекційний матеріал до теми 3, готуватись до практичних занять	20

1	2	3	4	5	6
Тема 4. Класифікація та кластеризація даних					
Лекція 5. Класифікація та кластеризація даних Задача класифікації. Задача кластеризації. Метрики оцінки. Дискримінантний аналіз. Логістична регресія. Метод k-найближчих сусідів. Відстань і схожість між об'єктами. Методи кластеризації. Кластеризація методом k-середніх (k-means).		Практична робота 8. Метод k-найближчих сусідів Практична робота 9. Логістична регресія Практична робота 10. Методи кластеризації Практична робота 11. Контрольна робота №1		опрацювати лекційний матеріал до теми 4, готуватись до практичних занять	22
Модуль 2. Інтелектуальний аналіз даних з використанням статистичних пакетів					
Тема 5. Інтелектуальний аналіз даних					
Лекція 6-7. Інтелектуальний аналіз даних Data Mining. Text Mining. Web Mining. Бібліотека scikit-learn. Векторизація тексту. Зменшення розмірності даних. Методи аналізу веб-даних. Частотний аналіз слів. Аналіз тональності тексту. Рекомендаційні системи.	2	Практична робота 12. Text Mining Практична робота 13. Web Mining Практична робота 14. Рекомендаційна система	2	опрацювати лекційний матеріал до теми 5, готуватись до практичних занять	18
Тема 6. Пошук асоціативних правил					
Лекція 8. Пошук асоціативних правил Основи задачі пошуку асоціативних правил: підтримка, достовірність, приріст. Алгоритм Apriori. Алгоритм FP-росту (FP-Growth).	2	Практична робота 15. Пошук асоціативних правил Практична робота 16. Контрольна робота №2		опрацювати лекційний матеріал до теми 6, готуватись до практичних занять	16
Всього, годин	8		4		108

Розділ 5. Система оцінювання знань студентів

Таблиця 5.1 - Розподіл балів за результатами вивчення навчальної дисципліни

Вид робіт	Максимальна кількість балів
-----------	-----------------------------

Вид робіт	Максимальна кількість балів
Модуль 1 . Основи аналізу даних та методи Data Mining	
<i>Практичні роботи 1-2 (тема 1)</i>	2 практичних роботи x 5 балів = 10
<i>Практичні роботи 3-4 (тема 2)</i>	2 практичних роботи x 5 балів = 10
<i>Практичні роботи 5-7 (тема 3)</i>	3 практичних роботи x 5 балів = 15
<i>Практичні роботи 8-10 (тема 4)</i>	3 практичних роботи x 5 балів = 15
<i>Практична робота 11</i> (Контрольна робота №1)	15
Всього за модулем 1	65
Модуль 2. Інтелектуальний аналіз даних з використанням статистичних пакетів	
<i>Практичні роботи 12-14 (тема 5)</i>	3 практичних роботи x 5 балів = 15
<i>Практична робота 15 (тема 6)</i>	1 практична робота x 5 балів = 5
<i>Практична робота 16</i> (Контрольна робота №2)	15
Всього за модулем 2	35
Всього по курсу	<u>100</u>

Система нарахування додаткових балів за видами робіт з вивчення навчальної дисципліни

Форма роботи	Вид роботи	Бали
Навчальна робота	Опрацювання додаткових навчальних курсів, написання та публічний захист наукового реферату, наукової статті на тему, що відповідає дисципліні	10*

За додаткові види навчальних робіт студент може отримати не більше 10 балів. Додаткові бали додаються до загальної підсумкової оцінки за вивчення навчальної дисципліни, але загальна підсумкова оцінка не може перевищувати 100 балів.

Таблиця 6 – Шкала оцінювання знань здобувачів вищої освіти за результатами вивчення навчальної дисципліни

Сума балів за всі види навчальної діяльності	Оцінка за шкалою ЄКТС	Оцінка за національною шкалою
90-100	A	Відмінно
82-89	B	Дуже добре
74-81	C	Добре
64-73	D	Задовільно
60-63	E	Задовільно достатньо
35-59	FX	Незадовільно з можливістю повторного складання
0-34	F	Незадовільно з обов'язковим повторним вивченням навчальної дисципліни

Розділ 6. Інформаційні джерела

1. https://prometheus.org.ua/course/course-v1:IRF+Stat101+2016_T3 Курс «Аналіз даних та статистичне виведення на мові R»

2. Saiz A.Z., Gonzalez C.Q., Gil L.H., Ruiz D.M. An Introduction to Data Analysis in R: Hands-on Coding, Data Mining, Visualization and Statistics from Scratch Springer, 2020. — 291 p.
3. Stanford University Data Mining Lecture Notes – 2020. – Режим доступу: <http://infolab.stanford.edu/~ullman/mining/2003.html>.
4. Інтелектуальний аналіз даних: практикум/ М.Т. Фісун, І.О. Кравець, П.П. Казмірчук. - Л.: Новий Світ-2000, 2020.- 162с. Режим доступу: <https://ns2000.com.ua/wp-content/uploads/2019/07/Intelektual-nyu-analiz-danykh-praktykum.pdf>
5. Литвин В.В. Аналіз даних та знань: підручник/ В.В. Литвин, В.В. Пасічник, Ю.В. Нікольський.- Л.:Магнолія, 2020.- 276с..
6. Фоусет Т., Провост Ф. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані – Київ:Наш Формат, 2019. – 400 с.
7. Шпігельгальтер Д. Мистецтво статистики. Прийняття аргументованих рішень на основі даних. /Девід Шпігельгальтер. Пер. із англ. Наталія Лавська. – К.: КМ-БУКС. 2023. – 384 с

Розділ 7. Програмне забезпечення навчальної дисципліни

Використовується наступне програмне забезпечення:

1. Персональний комп'ютер або ноутбук зі сталим доступом до мережі Інтернет.
2. Система дистанційного навчання ПУЕТ та його засоби відео-конференц зв'язку, дистанційний курс з ОК «Аналіз даних та прикладні пакети статистичної обробки».
3. Дистрибутив Anaconda та мова програмування Python 3.x.
4. MS Excel або інший табличний редактор.