

**ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД УКООПСПЛКИ
«ПОЛТАВСЬКИЙ УНІВЕРСИТЕТ ЕКОНОМІКИ І ТОРГІВЛІ»
Навчально-науковий інститут денної освіти
Кафедра комп'ютерних наук та інформаційних технологій**

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ О.В. Ольховська

« ____ » _____ 2021 р.

РОБОЧА ПРОГРАМА

навчальної дисципліни	«Аналіз даних та прикладні пакети статистичної обробки»
освітня програма	Комп'ютерні науки
спеціальність	122 Комп'ютерні науки
галузь знань	12 Інформаційні технології
ступінь вищої освіти	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 2 вересня 2021 року, №1

Полтава 2021

Укладач:

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

ПОГОДЖЕНО:

Гарант освітньої програми «Комп'ютерні науки» спеціальності 122
Комп'ютерні науки ступеня бакалавра, к.ф.-м.н, доцент

_____ О.О. Черненко

«___»_____2021 року

Зміст

робочої програми початкової дисципліни

<u>Розділ 1. Опис навчальної дисципліни</u>	4
<u>Розділ 2. Перелік компетентностей та програмні результати навчання</u>	4
<u>Розділ 3. Програма навчальної дисципліни</u>	5
<u>Розділ 4. Тематичний план вивчення навчальної дисципліни</u>	6
<u>Розділ 5. Оцінювання результатів навчання</u>	10
<u>Розділ 6. Інформаційні джерела</u>	11
<u>Розділ 7. Програмне забезпечення навчальної дисципліни</u>	12

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> Дипломне проектування	
Мова викладання	Українська	
Статус дисципліни	Вибіркова	
Курс/семестр вивчення	4/7	
Кількість кредитів ECTS/кількість модулів	4/2	
Денна форма навчання:		
Кількість годин: 120 год – загальна кількість: 1 семестр – 120 год.		
- Лекції: 16 год.		
- Практичні (семінарські, лабораторні) заняття: 32 год.		
- Самостійна робота: 72 год.		
- Вид підсумкового контролю (ПМК, екзамен): 1 семестр – ПМК		
Заочна форма навчання:		
Кількість годин: 120 год – загальна кількість: 1 семестр – 120 год.		
- Лекції: 4 год.		
- Практичні (семінарські, лабораторні) заняття: 2 год.		
- Самостійна робота: 114 год.		
- Вид підсумкового контролю (ПМК, екзамен): 1 семестр – ПМК		

Розділ 2. Перелік компетентностей та програмні результати навчання

Метою навчальної 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 «Інформатика»

<i>Компетентності, якими повинен оволодіти здобувач</i>	<i>Програмні результати навчання</i>
Загальні компетентності	
Здатність до абстрактного мислення, аналізу та синтезу (ЗК1).	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 комп'ютерних наук.
Здатність застосовувати знання у практичних ситуаціях (ЗК2).	
Знання та розуміння предметної області та розуміння професійної діяльності (ЗК3).	
Здатність спілкуватися державною мовою як усно, так і письмово (ЗК4).	ПР3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної
Здатність вчитися й оволодівати сучасними знаннями (ЗК6).	
Здатність до пошуку, оброблення та аналізу	

<p>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).</p> <p style="text-align: center;">Спеціальні компетентності</p> <p>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).</p>	<p>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..</p>
--	--

Розділ 3. Програма навчальної дисципліни

Семестр 1

Модуль 1. Основи аналізу даних. Прикладні пакети для статистичної обробки даних. Кореляційний та регресійний аналіз

Тема 1. Попередній аналіз

Підготовка вхідної інформації. Попередня обробка. Аномальні спостереження. Розвідувальний аналіз. Гіпотези та їх перевірка. Знайомство з Anaconda та Python. Видалення аномальних спостережень. Робота з даними в Python. Перевірка стохастичності вибірки. Основні статистичні характеристики.

Тема 2. Кореляційний аналіз

Кореляційний аналіз. Кореляція якісних змінних. Коефіцієнт кореляції. Кореляція якісних змінних. Частинна кореляція.

Тема 3. Регресійний аналіз

Регресійний аналіз. Парна лінійна регресія. Обчислення параметрів парної регресії засобами MS Excel. Парна не лінійна регресія. Множинна регресія.

Тема 4. Дисперсійний аналіз

Дисперсійний аналіз.

Тема 5. Дискримінантний аналіз

Дискримінантний аналіз. Метод k-найближчих сусідів.

Тема 6. Кластерний аналіз

Кластерний аналіз.

Модуль 2. Аналіз даних з використанням статистичних пакетів**Тема 7. Аналіз двовимірних даних**

Аналіз двовимірних даних

Тема 8. Класифікація даних

Класифікація даних

Розділ 4. Тематичний план вивчення навчальної дисципліни

Таблиця 3. 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 теорії програмування. Контекстовільні мови. Синтаксичний аналіз.					
Тема 1. Попередній аналіз <i>Лекція 1. Підготовка вхідної інформації. Попередня обробка. Аномальні спостереження Розвідувальний аналіз. Гіпотези та їх перевірка</i>	2	Практичне заняття 1. Знайомство з Anaconda та Python	2	опрацювати лекційний матеріал до теми 1, готуватись до практичних занять, опрацювати ДК	9
		Практичне заняття 2. Видалення аномальних спостережень	2		
		Практичне заняття 3. Робота з даними в Python	2		
Тема 2. Кореляційний аналіз <i>Лекція 2. Кореляційний аналіз. Кореляція якісних змінних</i>	2	Практичне заняття 4. Перевірка стохастичності вибірки	2	опрацювати лекційний матеріал до теми 2, готуватись до практичних занять, опрацювати ДК	9
		Практичне заняття 5. Основні статистичні характеристики	2		
Тема 3. Регресійний				опрацювати лекційний матеріал до	9

1	2	3	4	5	6
аналіз	2	Практичне заняття 6. Коефіцієнт кореляції	2	теми 3, готуватись до практичних занять, опрацювати ДК	9
<u>Лекція 3. Регресійний аналіз</u>		Практичне заняття 7. Кореляція якісних змінних	2		
		Практичне заняття 8. Частинна кореляція	2		
Тема 4. Дисперсійний аналіз	2	Практичне заняття 9. Парна лінійна регресія	2	опрацювати лекційний матеріал до теми 4, готуватись до практичних занять, опрацювати ДК	9
<u>Лекція 4. Дисперсійний аналіз</u>		Практичне заняття 10. Обчислення параметрів парної регресії засобами MS Excel	2		
Тема 5. Дискримінантний аналіз	2	Практичне заняття 11. Парна нелінійна регресія	2	опрацювати лекційний матеріал до теми 5, готуватись до практичних занять, опрацювати ДК	9
<u>Лекція 5. Дискримінантний аналіз</u>		Практичне заняття 12. Множинна регресія	2		
Тема 6. Кластерний аналіз	2	Практичне заняття 13. Модульна контрольна робота №1	2	опрацювати лекційний матеріал до теми 6, готуватись до практичних занять, опрацювати ДК, готуватись до модульної контрольної роботи №1	9
<u>Лекція 6. Кластерний аналіз</u>					
Модуль 2. Аналіз даних з використанням статистичних пакетів					
Тема 7. Аналіз двовимірних даних	2	Практичне заняття 14 Метод k-найближчих сусідів.	4	опрацювати лекційний матеріал до теми 7, готуватись до практичних занять, опрацювати ДК	9
<u>Лекція 7. Аналіз двовимірних даних</u>					
Тема 8. Класифікація даних	2	Практичне заняття 15 Модульна контрольна робота № 2.	2	опрацювати лекційний матеріал до	9
<u>Лекція 8. Класифікація</u>					

1	2	3	4	5	6
даних				теми 8, готуватись до практичних занять, опрацювати ДК, готуватись до модульної контрольної роботи №2	
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 теорії програмування. Контекстовільні мови. Синтаксичний аналіз.					
Тема 1. Попередній аналіз <i>Лекція 1. Підготовка вхідної інформації. Попередня обробка. Аномальні спостереження Розвідувальний аналіз. Гіпотези та їх перевірка</i>	2	<u>Практичне заняття 1.</u> Знайомство з Anaconda та Python <u>Практичне заняття 2.</u> Видалення аномальних спостережень <u>Практичне заняття 3.</u> Робота з даними в Python	2	опрацювати лекційний матеріал до теми 1, готуватись до практичних занять, опрацювати ДК	14
Тема 2. Кореляційний аналіз <i>Лекція 2. Кореляційний аналіз. Кореляція якісних змінних</i>	2	<u>Практичне заняття 4.</u> Перевірка стохастичності вибірки <u>Практичне заняття 5.</u> Основні статистичні характеристики		опрацювати лекційний матеріал до теми 2, готуватись до практичних занять, опрацювати ДК	14
Тема 3. Регресійний аналіз <i>Лекція 3. Регресійний аналіз</i>		<u>Практичне заняття 6.</u> Коефіцієнт кореляції <u>Практичне заняття 7.</u> Кореляція якісних змінних		опрацювати лекційний матеріал до теми 3, готуватись до практичних занять, опрацювати ДК	14

1	2	3	4	5	6	
<p>Тема 4. Дисперсійний аналіз</p> <p><u>Лекція 4. Дисперсійний аналіз</u></p>		<p><u>Практичне заняття 8.</u> Частинна кореляція</p>		<p>опрацювати лекційний матеріал до теми 4, готуватись до практичних занять, опрацювати ДК</p>	14	
		<p><u>Практичне заняття 9.</u> Парна лінійна регресія</p>			<p>опрацювати лекційний матеріал до теми 5, готуватись до практичних занять, опрацювати ДК</p>	14
<p>Тема 5. Дискримінантний аналіз</p> <p><u>Лекція 5. Дискримінантний аналіз</u></p>		<p><u>Практичне заняття 10.</u> Обчислення параметрів парної регресії засобами MS Excel</p>				<p>опрацювати лекційний матеріал до теми 6, готуватись до практичних занять, опрацювати ДК, готуватись до модульної контрольної роботи №1</p>
		<p><u>Практичне заняття 11.</u> Парна нелінійна регресія</p>		<p>опрацювати лекційний матеріал до теми 6, готуватись до практичних занять, опрацювати ДК, готуватись до модульної контрольної роботи №1</p>	14	
<p>Тема 6. Кластерний аналіз</p> <p><u>Лекція 6. Кластерний аналіз</u></p>		<p><u>Практичне заняття 12.</u> Множинна регресія</p>			<p>опрацювати лекційний матеріал до теми 6, готуватись до практичних занять, опрацювати ДК, готуватись до модульної контрольної роботи №1</p>	
		<p><u>Практичне заняття 13.</u> Модульна контрольна робота №1</p>				
Модуль 2. Аналіз даних з використанням статистичних пакетів						
<p>Тема 7. Аналіз двовимірних даних</p> <p><u>Лекція 7. Аналіз двовимірних даних</u></p>		<p><u>Практичне заняття 14</u> <u>Метод k-найближчих сусідів.</u></p>	2	<p>опрацювати лекційний матеріал до теми 7, готуватись до практичних занять, опрацювати ДК</p>	14	
<p>Тема 8. Класифікація даних</p> <p><u>Лекція 8. Класифікація даних</u></p>		<p><u>Практичне заняття 15</u> Модульна контрольна робота № 2.</p>		<p>опрацювати лекційний матеріал до теми 8, готуватись до практичних занять, опрацювати ДК, готуватись до модульної контрольної роботи</p>	16	

1	2	3	4	5	6
				№2	
Всього, годин	4		2		114

Розділ 5. Оцінювання результатів навчання

Таблиця 5. Розподіл балів за результатами вивчення навчальної дисципліни

Вид діяльності	Максимальна кількість балів за вид навчальної роботи
Модуль 1. Основи теорії програмування. Контекстовільні мови. Синтаксичний аналіз.	
Тема 1. Попередній аналіз	
Практичне заняття 1.	6
Практичне заняття 2.	6
Практичне заняття 3.	6
Тема 2. Кореляційний аналіз	
Практичне заняття 4.	6
Практичне заняття 5.	6
Тема 3. Регресійний аналіз	
Практичне заняття 6.	6
Практичне заняття 7.	6
Практичне заняття 8.	6
Тема 4. Дисперсійний аналіз	
Практичне заняття 9.	6
Практичне заняття 10.	6
Тема 5. Дискримінантний аналіз	
Практичне заняття 11.	6
Практичне заняття 12.	6
Тема 6. Кластерний аналіз	
Практичне заняття 13.	11
Модульна контрольна робота №1	
Всього за модулем 1	83
Модуль 2. Аналіз даних з використанням статистичних пакетів	
Тема 7. Аналіз двовимірних даних	
Практичне заняття 14.	6
Тема 8. Класифікація даних	
Практичне заняття 15.	6
Практичне заняття 15. Модульна контрольна робота № 2.	11
Всього за модулем 2	17
Всього по курсу	100

Розділ 6. Інформаційні джерела

Основні джерела

1. Фоусет Т., Провост Ф. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані – Київ:Наш Формат, 2019. – 400 с.
2. Маккинни У. Python и анализ данных. Второе издание. – Москва: ДМК Пресс, 2020. - 540с.
3. Вандер П. Дж. Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.: ил.
4. Брюс П., Практическая статистика для специалистов Data Science. 50 важнейших понятий. - СПб.: БХВ-Петербург, 2016. - 304с.
5. Силен Д., Мейсман А., Али М. Основы Data Science і Big Data. Python и наука о данных, СПб.: Питер Пресс, 2019. - 336 с.

Додаткові джерела

6. Джоэл Грас, Data Science. Наука про даних з нуля – СПб. : БХВ-Петербург, - 2020. - 416 с.
7. Jake VanderPlas. Python Data Science Handbook: Essential Tools for Working with Data 1st Edition. - O'reilly Media, 2016. - 548 с.

Розділ 7. Програмне забезпечення навчальної дисципліни

- Дистрибутив Anaconda та мова програмування Python 3.x.
- MS Excel або інший табличний редактор.
- Дистанційний курс з навчальної дисципліни «Аналіз даних та прикладні пакети статистичної обробки» в системі дистанційного навчання ПУЕТ