

КЛАСИЧНИЙ ПРИВАТНИЙ УНІВЕРСИТЕТ
ІНСТИТУТ ЕКОНОМІКИ ТА ПРАВА
КАФЕДРА ЕКОНОМІКИ ТА СТАТИСТИКИ

СИЛАБУС

навчальної дисципліни
«ПАКЕТИ АНАЛІЗУ ДАНИХ В PYTHON»

КОНТАКТНА ІНФОРМАЦІЯ ТА ТЕХНІЧНА ДОПОМОГА
(включаючи електронну пошту, робочий час / місцезнаходження тощо).

Викладач (-і)	Сергеева Людмила Нільсівна
Контактний тел.	+38(061)764-67-50 (внутр. 169)
E-mail:	lud.sergeeva@gmail.com
Сторінка курсу на сайті підтримки навчальних програм КПУ	http://www.zhu.edu.ua/cpu_edu/course/view.php?id=4554
Консультації	<i>Очні консультації:</i> за графіком консультацій викладача, а. 507, головний корпус КПУ <i>Консультації off-line:</i> шляхом повідомлення на сторінці навчальної дисципліни сайту підтримки навчальних програм КПУ http://www.zhu.edu.ua/cpu_edu/message

АНОТАЦІЯ

Великі обсяги даних, які характеризують різні економічні, соціальні, природничі процеси, для проведення ґрунтового аналізу вимагають застосування сучасних комп'ютерних технологій, зокрема вміння використовувати пакети, розроблені мовою програмування Python.

Освітній процес з дисципліни здійснюється за такими формами: навчальні заняття; самостійна робота; контрольні заходи. Видами навчальних занять згідно з навчальним планом є: лекції; практичні заняття, а також консультації.

Практичні заняття передбачають: перевірку засвоєння студентами лекційного матеріалу; виконання індивідуальних завдань з використання пакетів аналізу даних..

Самостійна робота студентів полягає у засвоєнні вивченого навчального матеріалу в час, вільний від обов'язкових навчальних занять, без участі викладача.

Консультації призначені для роз'яснення студентам теоретичних або практичних питань.

Засвоєння навчального матеріалу перевіряється за допомогою поточного контролю, який здійснюється на практичних заняттях у формі перевірки виконаних завдань, самостійних робіт.

Підсумковий (семестровий) контроль після завершення 1 семестру здійснюється у формі екзамену.

ФОРМАТ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Загальна кількість годин – 150 год., у т. ч. 56 годин аудиторних занять і 94 години самостійної роботи студента. Кількість кредитів ECTS – 5.

Всього кредитів	Всього годин	Аудиторних годин	У тому числі			Сам. робота
			Лекц.	Лабор.	Семін. (практ.)	
5	150	56	24	-	32	94

ОЗНАКИ ДИСЦИПЛІНИ

Навчальний рік	Курс (рік навчання)	Семестр	Цикл Підготовки	Обов'язкова / вибіркова
2024/2025	1	1	Професійна	обов'язкова

МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета навчальної дисципліни

Метою викладання навчальної дисципліни «Пакети аналізу даних в Python» є оволодіння сучасними методами, теоретичними положеннями та основними застосуваннями основних пакетів аналізу даних, розроблених мовою програмування Python.

Завдання навчальної дисципліни

формування здатності розв'язувати складні спеціалізовані математичні та статистичні задачі, що характеризується комплексністю і невизначеністю умов і передбачає застосування теоретико - ймовірнісних і статистичних алгоритмів; набуття знань, умінь та навичок (компетентностей) на рівні новітніх досягнень у математиці та статистиці.

ЗАПЛАНОВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ

У результаті вивчення дисципліни студенти повинні знати:

- як використовуються основні пакети аналізу даних для розв'язку певних задач аналізу статистичної інформації,
- обирати відповідний пакети і метод, інтерпретувати отримані результати з теоретико-ймовірнісної точки зору та змістовної якості,
- розуміти обмеження у використанні певних методів аналізу даних, й «пастки», котрих слід уникати.

Після вивчення дисципліни студенти повинні вміти:

- застосовувати ймовірнісно-статистичні моделі та методи;
- використовувати в практичній діяльності спеціалізоване програмне забезпечення;
- застосовувати пакети аналізу даних для розв'язання проблем.

Відповідно до освітньо-професійної програми підготовки магістра галузі знань 11 Математика та статистика за спеціальністю 112 Статистика, освітня програма Комп'ютерна статистика та аналіз даних вивчення дисципліни «Пакети аналізу даних в

Python» сприяє формуванню **компетентностей та програмних результатів навчання:**

Інтегральна компетентність

Здатність розв'язувати складні математичні та статистичні задачі, практичні проблеми у професійній діяльності або у процесі навчання, що передбачає проведення досліджень та здійснення інновацій і характеризується комплексністю та невизначеністю умов.

Загальні компетентності:

ЗК 3. Здатність вирішувати проблеми у професійній діяльності на основі абстрактного мислення, аналізу, синтезу та прогнозу.

ЗК 4. Здатність до пошуку, оброблення й аналізу інформації з різних джерел, необхідної для розв'язування наукових і професійних завдань.

ЗК 7. Здатність до виконання дослідницької роботи з елементами наукової новизни.

Спеціальні (фахові, предметні) компетентності:

СК 1. Знання на рівні новітніх досягнень, необхідні для дослідницької та інноваційної діяльності у сфері статистики та її практичних застосувань

СК 10. Здатність до самоосвіти та підвищення кваліфікації на основі інноваційних підходів у сфері статистики.

СК 12. Володіння знаннями та здатність ініціювати й проводити наукові дослідження у спеціалізованих областях статистики.

Програмні результати навчання:

РН 1. Знати та розуміти фундаментальні і прикладні аспекти наук у сфері статистики.

РН 2. Відтворювати знання фундаментальних розділів статистики в обсязі, необхідному для володіння математичним та економічним апаратами відповідної галузі знань і використання статистичних методів у обраній професії.

РН 4. Володіти знаннями грамотної побудови комунікації в освітньому і науковому процесі, відбору вихідних даних дослідження, складання списку використаних джерел, опису наукових результатів.

РН 5. Уміти використовувати фундаментальні закономірності статистики у професійній діяльності.

РН 8. Ініціювати і проводити наукові дослідження у спеціалізованій області математичної статистики та розв'язувати задачі в інших галузях знань методами математичного та економічного моделювання.

РН 9. Інтегрувати знання з різних галузей для вирішення теоретичних та практичних задач і проблем.

РН 10. Застосовувати нові підходи для вироблення стратегії прийняття рішень у складних непередбачуваних умовах.

РН 12. Уміти самостійно планувати виконання дослідницького та інноваційного завдання та формулювати висновки за його результатами.

РН 13. Усно й письмово спілкуватися рідною та іноземними мовами в науковій, виробничій та соціально-суспільній сферах діяльності із професійних питань; читати спеціальну літературу; знаходити, аналізувати та використовувати інформацію з різних довідкових джерел.

РН 14. Використовувати раціональні способи пошуку та використання науково-технічної інформації в галузі статистики, включаючи засоби електронних інформаційних мереж; застосовувати інформаційні ресурси, у тому числі електронні, для пошуку відповідних математичних моделей.

ПЛАН КУРСУ

Назва змістових модулів та тем	Лекц.	Прак т (сем.)	Завдання для самостійної роботи
ЗМІСТОВИЙ МОДУЛЬ № 1. ТЕОРЕТИЧНІ ОСНОВИ ПРОВЕДЕННЯ СТАТИСТИЧНОГО АНАЛІЗУ ДАНИХ З ВИКОРИСТАННЯМ ПАКЕТІВ В PYTHON			
Тема 1. Вступ до аналізу даних в Python	2	2	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Основні етапи аналізу даних 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури.
Тема 2. Типи даних в Python. Імпорт даних з файлів різних форматів: формати .xlsx та CSV	2	2	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Різниця між форматами в Numpy та Pandas. 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання.
Тема 3. Імпорт даних з файлів різних форматів. Формати JSON, HTML та SQL	2	2	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Взаємодія з базами даних. 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури.
Тема 4. Математичні основи аналізу даних: теорія ймовірностей та математична статистика	2	2	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Порівняння Adaline та логістичної регресії. 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання.
Тема 5. Схема тестування статистичних гіпотез	2	2	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Характеристика та аналіз статистичних помилок. 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання.
Тема 6. Аналіз зв'язку (кореляція та асоціація).	2	2	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Методи візуалізації в аналізі зв'язку. 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання.
Тема 7. Опис якісних ознак	2	2	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Порівняння результатів проведення двох тестів. 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання.
Тема 8. Регресійний аналіз	2	2	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Логістична регресія в Python. 3. Підготовка практичного заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання
ЗМІСТОВИЙ МОДУЛЬ № 2. ОГЛЯД МОЖЛИВОСТЕЙ ПАКЕТІВ SCIPY.STATS, PANDAS ТА STATMODELS			
Тема 9. Особливості застосування пакету	2	4	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань:

Назва змістових модулів та тем	Лекц.	Прак т (сем.)	Завдання для самостійної роботи
scipy.stats в аналізі даних			1) Методи описової статистики в scipy.stats. 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання.
Тема 10. Особливості застосування пакету pandas в аналізі даних	2	4	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Робота з часовими рядами в pandas 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання.
Тема 11. Особливості застосування пакету statmodels: регресійний аналіз та аналіз часових рядів	2	4	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Переваги роботи з часовими рядами в statmodels у порівнянні з pandas. 3. Підготовка до семінарського заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання.
Тема 12. Особливості застосування пакету statmodels: аналіз виживаності	2	4	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: 1) Характеристика особливостей даних для аналізу виживаності 3. Підготовка до практичного заняття, опрацювання першоджерел та навчальної літератури; виконання індивідуального завдання.

ФОРМИ КОНТРОЛЮ ТА КРИТЕРІЇ ОЦІНЮВАННЯ

У процесі вивчення навчальної дисципліни «Пакети аналізу даних в Python» використовуються наступні види контролю:

1. Поточний контроль – здійснюється протягом семестру шляхом перевірки виконання завдань практичних занять, виконання індивідуальних завдань тощо. За змістом він включає перевірку ступеню засвоєння студентом навчального матеріалу, який охоплюється темою лекційного та практичного заняття, уміння самостійно опрацьовувати навчально-методичну літературу, здатність осмислювати зміст теми, уміння використовувати програмне забезпечення та здійснювати інтерпретацію отриманих результатів.

2. Підсумковий семестровий контроль – здійснюється у формі екзамену відповідно до графіка освітнього процесу.

Для оцінювання студентів використовується система накопичування балів. Згідно з «Положенням про організацію освітнього процесу в КПУ» підсумкова оцінка з дисципліни виставляється за 100-бальною шкалою з наступним переведенням у національну шкалу та шкалу ECTS.

Бали нараховуються за виконання завдань практичних робіт, індивідуальних завдань.

Результати поточного контролю здобувачів вищої освіти є складовими елементами підсумкової оцінки з дисципліни.

Оцінка рівня роботи студента протягом семестру під час навчальних занять та самостійної роботи здійснюється у межах 60 балів. Вага екзамену у підсумковій оцінці складає 40 балів.

Для оцінювання результатів навчання студентів з навчальної дисципліни «Пакети аналізу даних в Python» використовуються наступні критерії:

– рівню «відмінно» відповідає повністю виконане завдання, при якому студент

показав всебічне системне знання матеріалу; засвоєння літератури; чітке володіння понятійним апаратом, методами, методиками та інструментами використання пакетів аналізу даних; уміння надавати змістовну інтерпретацію отриманих результатів;

– рівню «добре» відповідає успішне виконання запропонованого завдання і засвоєння матеріалу літератури без змістовної інтерпретації отриманих результатів;

– рівню «задовільно» відповідає часткове виконання запропонованого завдання і засвоєння матеріалу літератури, при якому студент показав достатній рівень знань, але не зміг переконливо аргументувати свою відповідь, помилився у використанні понятійного апарату, показав недостатні знання рекомендованої літератури;

– рівню «незадовільно» відповідає відсутність виконання запропонованого завдання; студент показав незадовільні знання понятійного апарату і літератури чи взагалі нічого не відповів.

РОЗПОДІЛ БАЛІВ ПОТОЧНОГО ТА ПІДСУМКОВОГО (СЕМЕСТРОВОГО) ОЦІНЮВАННЯ

Контрольний захід	Бали
Модуль №1	30
Практичні роботи	20
Поточний контроль	10
Модуль №2	30
Практичні роботи	20
Поточний контроль	10
Екзамен	40
Разом	100

Шкала оцінювання: 100-бальна, національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка за шкалою ЄКТС	Оцінка за національною шкалою для екзамену
90-100	A	відмінно
82-89	B	добре
75-81	C	
67-74	D	
60-66	E	задовільно
35-59	FX	незадовільно з можливістю повторного складання
0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних / Я. В. Іванчук, В. І. Месюра, А. А. Яровий, О. Д. Манжілевський – Вінниця : ВНТУ, 2021. – 69 с.

2. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. – Київ. – 2017. – 150 с.

3. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. – К.: ІСЗІ КПІ ім. Ігоря Сікорського», 2018. — 297 с.

4. Wes McKinney. Python for Data Analysis. // <https://bedford-computing.co.uk/learning/wpcontent/uploads/2015/10/Python-for-Data-Analysis.pdf>

5. VanderPlas J. Python Data Science Handbook. – <https://jakevdp.github.io/PythonDataScienceHandbook/>.

6. Practical Statistics for Data Scientists / P. Bruce, A. Bruce, P. Gedeck. – O'Reilly Media, 2020.

ІНФОРМАЦІЙНІ РЕСУРСИ В ІНТЕРНЕТІ

1. UCI machine learning repository. <https://archive.ics.uci.edu/>
2. Kaggle: Your Machine Learning and Data Science Community <https://www.kaggle.com/>