

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

**СИЛАБУС**  
 навчальної дисципліни  
**«Аналіз даних та прикладні пакети статистичної  
 обробки»**  
 на 2024-2025 навчальний рік

Курс та семестр вивчення	4 курс, 7 семестр
Освітня програма/спеціалізація	122 Комп'ютерні науки
Спеціальність	122 Комп'ютерні науки
Галузь знань	12 «Інформаційні технології»
Ступінь вищої освіти	бакалавр

ПІБ НПП, який веде дану дисципліну,  
 науковий ступінь і вчене звання,  
 посада

Олексійчук Ю.Ф., к.ф.-м.н., доцент кафедри КНІТ  
 Оріхівська О. Г., старший викладач кафедри КНІТ

Контактний телефон	+380953000324
Електронна адреса	olexijchuk@gmail.com orikhivska@ukr.net
Розклад навчальних занять	<a href="http://schedule.puet.edu.ua/">http://schedule.puet.edu.ua/</a>
Консультації	Онлайн: субота о 8-00, згідно розкладу консультацій он-лайн: електронною поштою
Сторінка дистанційного курсу	<a href="https://el.puet.edu.ua/">https://el.puet.edu.ua/</a>

**Опис навчальної дисципліни**

<b>Мета вивчення навчальної дисципліни</b>	Основною 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.
<b>Тривалість</b>	4 кредити ЄКТС/120 годин (лекції 16 год., лабораторні заняття 32 год., самостійна робота 72 год.)
<b>Форми та методи навчання</b>	Лекції та практичні заняття в аудиторії, самостійна робота поза розкладом Методи навчання: інтерактивного навчання, наочний, словесні
<b>Система поточного та підсумкового контролю</b>	Поточний контроль: поточна модульна робота Підсумковий контроль: ПМК
<b>Базові знання</b>	Курс базується на таких дисциплінах: Теорія ймовірностей і математична статистика
<b>Мова викладання</b>	Українська

**Перелік компетентностей, які забезпечує дана навчальна дисципліна,  
 програмні результати навчання**

<b>Програмні результати навчання</b>	<b>Компетентності, якими повинен оволодіти здобувач</b>
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. ПР3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей	Здатність до абстрактного мислення, аналізу та синтезу (ЗК1). 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).

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.

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).

### Тематичний план навчальної дисципліни

Назва теми	Види робіт	Завдання самостійної роботи у розрізі тем
<b>Модуль 1. Основи аналізу даних та методи Data Mining</b>		
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.	відвідування занять; опитування на заняттях; опитування в процесі індивідуально-консультативних занять для перевірки засвоєння матеріалу пропущених занять.	опрацювати теорет. матеріал до теми 1, готуватись до практичних занять з теми
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).	відвідування занять; опитування на заняттях; опитування в процесі індивідуально-консультативних занять для перевірки засвоєння матеріалу пропущених занять.	опрацювати теорет. матеріал до теми 2, готуватись до практичних занять з теми
Тема 3. Регресійний аналіз  Проста лінійна регресія. Множинна лінійна регресія. Нелінійна регресія. Проблема мультиколінеарності. Перевірка якості моделі. Прогнозування.	відвідування занять; опитування на заняттях; опитування в процесі індивідуально-консультативних занять для перевірки засвоєння матеріалу пропущених занять.	опрацювати теорет. матеріал до теми 3, готуватись до практичних занять з теми

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).	відвідування занять; опитування на заняттях; опитування в процесі індивідуально-консультативних занять для перевірки засвоєння матеріалу пропущених занять; виконання модульних контрольних робіт.	опрацювати теорет. матеріал до теми 4, готуватись до практичних занять з теми та контрольної роботи
<b>Модуль 2. Інтелектуальний аналіз даних з використанням статистичних пакетів</b>		
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.	відвідування занять; опитування на заняттях; опитування в процесі індивідуально-консультативних занять для перевірки засвоєння матеріалу пропущених занять	опрацювати теорет. матеріал до теми 5, готуватись до практичних занять з теми
Тема 6. Пошук асоціативних правил Основи задачі пошуку асоціативних правил: підтримка, достовірність, приріст. Алгоритм Apriori. Алгоритм FP-росту (FP-Growth).	відвідування занять; опитування на заняттях; опитування в процесі індивідуально-консультативних занять для перевірки засвоєння матеріалу пропущених занять; перевірка виконання модульних контрольних робіт.	опрацювати теорет. матеріал до теми 6, готуватись до практичних занять з теми та контрольної роботи

### **Інформаційні джерела**

1. [https://prometheus.org.ua/course/course-v1:IRF+Stat101+2016\\_T3](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

### **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.

### **Політика вивчення навчальної дисципліни та оцінювання**

**Політика оцінювання здобувачів вищої освіти.** Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку (75% від можливої максимальної кількості балів за вид діяльності). Перескладання модулів відбувається із дозволу провідного викладача за наявності поважних причин (наприклад, лікарняний).

[Положення про організацію освітнього процесу](#)

[Положення про порядок та критерії оцінювання знань, вмінь та навичок здобувачів вищої освіти](#)

[Порядок ліквідації здобувачами вищої освіти академічної заборгованості](#)

**Політика щодо відвідування.** Відвідування занять є обов'язковим компонентом. За об'єктивних причин (наприклад, хвороба, працевлаштування, міжнародне стажування) навчання може відбуватись в режимі он-лайн.

**Політика щодо академічної доброчесності.** Здобувач повинен дотримуватися принципів академічної доброчесності, зокрема недопущення академічного плагіату, фальсифікації, фабрикації, списування під час поточного, рубіжного та підсумкового контролю. Списування під час контрольних робіт та поточних тестів заборонені (в т.ч. із використанням мобільних девайсів). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування та підготовки практичних завдань в процесі заняття. В ПУЕТ діють:

[Кодекс честі студента](#)

[Положення про академічну доброчесність](#)

[Положення про запобігання випадків академічного плагіату](#)

**Політика визнання результатів навчання визначена такими документами:**

[Положення про порядок перезарахування результатів навчання, здобутих в іноземних та вітчизняних закладах освіти](#)

[Положення про академічну мобільність здобувачів вищої освіти](#)

[Положення про порядок визнання результатів навчання здобутих шляхом неформальної та/або інформальної освіти; інфографіка](#) (розділ Освіта/Організація освітнього процесу/Неформальна освіта)

**Політика вирішення конфліктних ситуацій:**

[Положення про правила вирішення конфліктних ситуацій](#)

[Положення про апеляцію результатів підсумкового контролю у формі екзамену](#)  
уповноважена особа з питань запобігання та виявлення корупції

**Політика підтримки учасників освітнього процесу:**

[Психологічна служба](#)

[Студентський омбудсмен \(Уповноважений з прав студентів\) ПУЕТ](#)

[Уповноважений з прав корупції](#)

**Безпека освітнього середовища:** [Інформація про безпечність освітнього середовища ПУЕТ наведена у вкладці «Безпека життєдіяльності»](#)

## Оцінювання

Підсумкова оцінка за вивчення навчальної дисципліни розраховується через поточне оцінювання

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

Система нарахування додаткових балів за видами робіт з вивчення навчальної дисципліни

Форма роботи	Вид роботи	Бали
<b>Навчальна робота</b>	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.

**Шкала оцінювання здобувачів вищої освіти за результатами вивчення навчальної дисципліни**

<b>Сума балів за всі види навчальної діяльності</b>	<b>Оцінка за шкалою ЄКТС</b>	<b>Оцінка за національною шкалою</b>
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