КЛАСИЧНИЙ ПРИВАТНИЙ УНІВЕРСИТЕТ

Кафедра економіки та статистики

СИЛАБУС

навчальної дисципліни

ПП 6. Статистичні алгоритми навчання

2026-2027 навчальний рік

Викладач	Дяченко Олександр Григорович, кандидат економічних наук, доцент
Контактний телефон	+38097-333-01-12
Ідентифікатор та код доступу (Zoom / Google Meet)	318-745-5856, код доступу 8M5PeX
E-mail:	dyachenko09@meta.ua
Сторінка курсу на сайті підтримки навчальних програм КПУ	http://www.zhu.edu.ua/cpu_edu/course/view.php?id=4682
Консультації	Консультації проводяться в дистанційному режимі за попереднім погодженням із викладачем через месенджери або електронну пошту

ІНФОРМАЦІЯ ПРО НАВЧАЛЬНУ ДИСЦИПЛІНУ

Рівень вищої освіти	Другий (магістерський)	
Галузь знань	Е Природничі науки, математика та статистика	
Спеціальність	Е8 Статистика	
Освітня програма	Комп'ютерна статистика та аналіз даних	
Форма навчання	Очна (денна), заочна	
Мова викладання	Українська	
Курс (рік навчання), семестр	Курс 1, семестр 2	
Цикл підготовки	Професійної підготовки	
Статус дисципліни	Обов’язкова	
Обсяг дисципліни	Кількість кредитів ECTS	5
	Загальна кількість годин	150
	Аудиторних годин	60
	Лекції	24
	Практичні	36
	Самостійна робота	90
Форма контролю	Екзамен	
Попередні компоненти (Пререквізити)	ПП 1. Математико-статистичні методи і моделі в прийнятті управлінських рішень, ПП 4. Пакети аналізу даних у Python	
Паралельні та подальші компоненти (Постреквізити)	Паралельні: ПП 5. Імітаційне моделювання Подальші: П 1 Переддипломна практика, А 1 Підготовка та захист кваліфікаційної роботи магістра.	

МЕТА, ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета навчальної дисципліни

Формування здатності обґрунтовано добирати, програмно реалізовувати та оцінювати статистичні алгоритми навчання для розв'язання прикладних задач класифікації, регресії й кластеризації; перевіряти якість моделей на нових даних, інтерпретувати результати та забезпечувати відтворюваність обчислень.

Завдання навчальної дисципліни:

- опанування статистичних засад навчання моделей, функцій втрат, оптимізації та оцінювання узагальнювальної здатності;
- формування вмінь готувати дані до моделювання, обґрунтовувати способи їх поділу та запобігати витоку інформації;
- набуття навичок застосування алгоритмів класифікації, регресії та кластеризації;
- опанування методів регуляризації, перехресної перевірки й налаштування гіперпараметрів;
- розвиток умінь порівнювати моделі за обґрунтовано обраними показниками якості та перевіряти стійкість результатів;
- формування навичок програмної реалізації алгоритмів засобами Python і відкритих бібліотек;
- набуття досвіду виконання прикладного дослідження на реальних даних, документування обчислень і представлення результатів.

У результаті вивчення дисципліни здобувачі вищої освіти повинні знати:

- основні постановки задач навчання з учителем, без учителя та з підкріпленням;
- статистичні засади навчання моделей, функції втрат, емпіричний ризик і причини перенавчання;
- принципи поділу даних на навчальну, валідаційну й тестову частини;
- методи попереднього опрацювання даних та запобігання витоку інформації;
- призначення, припущення й обмеження основних алгоритмів класифікації, регресії та кластеризації;
- принципи регуляризації, перехресної перевірки та добору гіперпараметрів;
- показники якості моделей та умови їх застосування;
- підходи до інтерпретації моделей і пояснення отриманих результатів;
- вимоги до документування даних, програмного середовища та обчислювальних процедур.

Після вивчення дисципліни здобувачі вищої освіти повинні вміти:

- формулювати прикладну задачу та визначати тип статистичного навчання;
- перевіряти якість даних, виявляти пропуски, дублікати, аномальні значення й дисбаланс класів;
- обирати схему поділу даних відповідно до їх структури та мети дослідження;
- створювати послідовності попереднього опрацювання й навчання моделей без витоку даних;
- реалізовувати та застосовувати основні алгоритми класифікації, регресії й кластеризації;
- добирати функції втрат і показники якості відповідно до змісту задачі;
- виконувати перехресну перевірку та налаштування гіперпараметрів без використання тестової вибірки для добору моделі;
- порівнювати результати з базовою моделлю, аналізувати помилки та ознаки перенавчання;
- реалізовувати базовий алгоритм навчання з підкріпленням на навчальному

прикладі та пояснювати отримані результати;

- інтерпретувати результати, пояснювати обмеження й можливі наслідки застосування моделі;
- готувати відтворюваний програмний звіт і представляти результати прикладного дослідження.

КОМПЕТЕНТНОСТІ ТА ПРОГРАМНІ РЕЗУЛЬТАТИ НАВЧАННЯ

Інтегральна компетентність

Здатність розв'язувати складні, зокрема міждисциплінарні, задачі та проблеми у сфері статистики й комп'ютерного аналізу даних шляхом інтегрування сучасних знань, проведення досліджень та/або інноваційної діяльності й ухвалення обґрунтованих рішень за умов невизначеності та неповної інформації.

Загальні компетентності

ЗК 1. Здатність до безперервного навчання, професійної рефлексії, самостійного оволодіння сучасними знаннями та адаптації до змін у професійній сфері.

ЗК 2. Здатність інтегрувати знання з математичних, природничих, соціально-гуманітарних та економічних наук під час розв'язання складних професійних і дослідницьких задач у сфері статистики.

ЗК 3. Здатність до абстрактного, критичного та системного мислення, аналізу й синтезу під час розв'язання складних професійних і дослідницьких задач.

Спеціальні (фахові) компетентності

СК 1. Здатність застосовувати поглиблені сучасні знання зі статистики та комп'ютерного аналізу даних для розв'язання складних професійних і дослідницьких задач.

СК 2. Здатність інтегрувати статистичні, економетричні та обчислювальні методи для дослідження складних міждисциплінарних проблем.

СК 5. Здатність будувати, досліджувати й оцінювати математичні та ймовірнісно-статистичні моделі реальних процесів і систем.

СК 7. Здатність добирати, адаптувати та програмно реалізовувати статистичні методи й обчислювальні алгоритми для розв'язання складних прикладних задач.

Програмні результати навчання

РН 1. Застосовувати поглиблені сучасні знання зі статистики та комп'ютерного аналізу даних для розв'язання складних професійних і дослідницьких задач.

РН 3. Будувати, досліджувати й оцінювати математичні та ймовірнісно-статистичні моделі стохастичних, природничих і соціально-економічних процесів.

РН 5. Застосовувати статистичні методи для опрацювання, аналізу та інтерпретації даних під час розв'язання прикладних задач.

РН 9. Інтегрувати знання зі статистики, математики, економіки та відповідних предметних галузей для розв'язання складних теоретичних і прикладних задач.

РН 14. Застосовувати програмні засоби для підготовки, перевірки якості, опрацювання, аналізу та візуалізації даних, програмно реалізовувати статистичні методи й обчислювальні алгоритми для розв'язання прикладних задач.

ЗМІСТ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

РОЗДІЛ 1. СТАТИСТИЧНІ ЗАСАДИ НАВЧАННЯ ТА ПІДГОТОВКА ДАНИХ
Тема 1. Постановка задач статистичного навчання.
Тема 2. Підготовка даних до навчання моделей.
Тема 3. Оцінювання якості та вибір моделі.
Тема 4. Лінійні алгоритми навчання та оптимізація.
РОЗДІЛ 2. КЛАСИФІКАЦІЯ ТА РЕГРЕСІЯ
Тема 5. Логістична регресія та метод опорних векторів.
Тема 6. Метод найближчих сусідів і дерева рішень.
Тема 7. Ансамблеві методи навчання.
Тема 8. Регресійні моделі та регуляризація.
Тема 9. Показники якості та аналіз помилок моделей.
РОЗДІЛ 3. НАВЧАННЯ БЕЗ УЧИТЕЛЯ, НАВЧАННЯ З ПІДКРІПЛЕННЯМ ТА ПРИКЛАДНЕ ДОСЛІДЖЕННЯ
Тема 10. Кластерний аналіз.
Тема 11. Зменшення розмірності та візуалізація даних.
Тема 12. Основи навчання з підкріпленням.
Тема 13. Відтворюване прикладне дослідження та представлення результатів.

ФОРМИ КОНТРОЛЮ ТА ОЦІНЮВАННЯ

Контроль результатів навчання та рівня сформованості компетентностей здобувачів здійснюється на засадах прозорості та академічної доброчесності відповідно до Положення про критерії оцінювання результатів навчання здобувачів вищої освіти у Класичному приватному університеті. http://virtuni.education.zp.ua/info_cpu/node/5478

У процесі вивчення навчальної дисципліни використовуються такі види контролю:

1. **Поточний контроль:** охоплює виконання практичних завдань із програмної реалізації та застосування статистичних алгоритмів навчання, тематичне тестування, підготовку й представлення результатів прикладного дослідження.

2. **Семестровий (підсумковий) контроль:** проводиться у формі **екзамену** шляхом тестування за матеріалом усіх тем навчальної дисципліни.

Кожна практична робота, самостійна робота, тематичний тест та екзаменаційний тест оцінюються за 100-бальною шкалою. Підсумкова оцінка з дисципліни визначається як середньозважена величина з урахуванням коефіцієнтів, наведених у таблиці. Для тематичного тестування використовується середня арифметична оцінок за тематичні тести. Підсумкова оцінка переводиться в національну шкалу та шкалу ЄКТС.

Підсумковий бал = ($B_1 \times k_1$) + ($B_2 \times k_2$) +… +($B_n \times k_n$) + ($B_{екз.} \times k_{екз.}$)

де:

B_1, B_2, B_3,...., B_n – оцінки за практичні роботи, самостійну роботу та тематичне тестування, визначені за 100-бальною шкалою; k_1, k_2, k_3,...., k_n – відповідні вагові коефіцієнти.;

$B_{екз.}$ – бал за підсумковий семестровий контроль (екзамен), виставлений за 100-бальною шкалою; $k_{екз.}$ – ваговий коефіцієнт підсумкового семестрового контролю (екзамену).

Сума вагових коефіцієнтів поточного та підсумкового контролю дорівнює 1

$$k_1 + k_2 + k_3 + ... + k_n + k_{екз.} = 1$$

Структура компонентів оцінювання та їх питома вага

Види навчальної діяльності здобувача	Шкала оцінювання завдання	Ваговий коеф.	Питома вага у загальній оцінці
Практична робота 1. Підготовка даних до моделювання: перевірка якості, поділ вибірки та запобігання витоку інформації	100	0,10	10
Практична робота 2. Програмна реалізація лінійного алгоритму навчання та дослідження його збіжності	100	0,05	5
Практична робота 3. Порівняння класифікаторів: логістична регресія, метод опорних векторів і метод найближчих сусідів	100	0,10	10
Практична робота 4. Побудова дерев рішень та ансамблевих моделей, налаштування параметрів і порівняння якості	100	0,10	10
Практична робота 5. Регресійні моделі: регуляризація та перевірка прогнозної якості	100	0,10	10
Практична робота 6. Кластеризація, зменшення розмірності та інтерпретація результатів	100	0,10	10
Практична робота 7. Реалізація алгоритму навчання з підкріпленням на навчальному прикладі	100	0,05	5
Самостійна робота. Аналіз реальних даних із застосуванням алгоритмів навчання: програмний код, візуалізація результатів і звіт із висновками.	100	0,15	15
Підсумкове тестування	100	0,05	5
Екзамен	100	0,20	20
Разом	**100**	1	**100**
Додаткові досягнення з навчальної дисципліни (за наявності) Результати науково-дослідної діяльності здобувача, зокрема за такі активності: – участь у предметних олімпіадах, конкурсах студентських наукових робіт з навчальної дисципліни; – участь у засіданнях наукових гуртків, зокрема в обговореннях у форматі круглих столів, присвячених актуальним темам з навчальної дисципліни; – публікації результатів досліджень у вигляді наукових статей, тез доповідей, а також виступи на наукових конференціях.	**До 10 балів**		

КРИТЕРІЇ ОЦІНЮВАННЯ

100-бальна система	шкала ЄКТС	Національна шкала	Характеристика вимог та критеріїв оцінювання
90–100	A	Відмінно	Здобувач самостійно формулює задачу, обґрунтовує вибір алгоритмів, коректно готує й розділяє дані. Реалізує моделі, налаштовує параметри без використання тестових даних для добору моделі, порівнює результати з базовим підходом. Аналізує помилки й обмеження, забезпечує відтворюваність обчислень та аргументує висновки.
82–89	B	Добре	Правильно реалізує основні алгоритми, обирає показники якості та процедуру перевірки. Результати обґрунтовані й задокументовані. Окремі неточності не змінюють основних висновків.
75–81	C	Добре	Загалом правильно виконує типові задачі, застосовує належний поділ даних і порівнює моделі. Допускає окремі недоліки в налаштуванні, інтерпретації або документуванні, які потребують уточнення.
67–74	D	Задовільно	Виконує типові завдання за відомим алгоритмом, але неповно обґрунтовує вибір моделі й показників якості. Має труднощі з аналізом перенавчання, помилок та обмежень результатів.
60–66	E	Задовільно	Демонструє мінімально достатнє розуміння основних алгоритмів. За інструкцією навчає модель і визначає показники якості, проте потребує допомоги під час вибору процедури перевірки та пояснення результатів.
35–59	FX	Незадовільно	Допускає суттєві помилки в підготовці даних, реалізації алгоритму або оцінюванні якості, через що висновки не мають достатнього обґрунтування. Мінімальний рівень результатів навчання не досягнутий. Потрібне повторне складання.
0–34	F	Незадовільно	Не розуміє основних принципів статистичного навчання, не може виконати базове завдання або пояснити процедуру перевірки якості моделі. Потрібне повторне вивчення дисципліни.

ШКАЛА ОЦІНЮВАННЯ: 100-БАЛЬНА, НАЦІОНАЛЬНА ТА ECTS

Сума балів за всі види навчальної діяльності	Оцінка за шкалою ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проєкту (роботи), практики	для заліку
90-100	A	відмінно	зараховано
82-89	B	добре	
75-81	C		
67-74	D	задовільно	
60-66	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	F	незадовільно з обов’язковим повторним вивченням дисципліни	не зараховано з обов’язковим повторним вивченням дисципліни

ПОЛІТИКА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Політика щодо термінів виконання завдань та перескладання	З огляду на безпекову ситуацію, повітряні тривоги та перебої з електропостачанням, освітній процес базується на принципах гнучкості та академічної підтримки. Усі завдання виконуються здобувачами в асинхронному режимі без застосування штрафних санкцій чи зниження балів за порушення первинних термінів, якщо затримка зумовлена об'єктивними обставинами непереборної сили. Здавання та перескладання пропущених завдань чи тем здійснюється у зручному для студента темпі протягом семестру за погодженням із викладачем без будь-яких обмежень.
Політика академічної доброчесності	Здобувачі освіти зобов'язані неухильно дотримуватись принципів академічної доброчесності під час виконання усіх видів навчальної, наукової та творчої діяльності. Порушеннями академічної доброчесності вважаються: академічний плагіат, самоплагіат, фабрикація, фальсифікація, списування, обман, хабарництво, необ'єктивне оцінювання та академічне шахрайство. Здобувачам освіти забороняється вдаватися до будь-яких форм академічного шахрайства, зокрема: використовувати під час контрольних заходів заборонені матеріали або технічні засоби; посилатися на джерела, які не використовувались у роботі; здавати роботи з однаковим змістом, підготовлені різними особами. У разі встановлення порушення академічної доброчесності до здобувача освіти можуть бути застосовані такі заходи академічної відповідальності, як повторне проходження оцінювання (контрольна робота, екзамен, залік тощо), повторне вивчення відповідного освітнього компонента або відрахування з університету.
Політика щодо відвідування занять в умовах воєнного стану	Відвідування навчальних занять здобувачами вищої освіти є обов'язковим відповідно до затвердженого розкладу. *Організація освітнього процесу в умовах загрози та повітряних тривог* У разі включення сигналу «Повітряна тривога» під час проведення занять у режимі реального часу (Zoom / Google Meet) або перебування в приміщеннях Університету, проведення синхронного заняття негайно припиняється, а всі учасники освітнього процесу зобов'язані перейти до укриттів. Навчальне заняття, перерване через сигнал «Повітряна тривога» або яке не відбулося через її тривалість, автоматично переводиться в асинхронний режим. Відсутність здобувача освіти на синхронній конференції під час повітряної тривоги, відсутності електроенергії або зв'язку не є підставою для виставлення заборгованості чи зниження підсумкової оцінки, за умови опрацювання навчального матеріалу та виконання завдань в асинхронному режимі. *Застосування дистанційних технологій та форс-мажорних обставин* У разі переходу університету або його окремих структурних підрозділів на навчання із використанням дистанційних технологій через безпекову ситуацію чи інші форс-мажорні обставини, освітній процес регулюється *«Положенням про організацію освітнього процесу здобувачів вищої освіти з використанням дистанційних технологій при обмежені або відсутності фізичного відвідування закладів освіти здобувачами з причин непереборної сили (природні катаклізми, заходи карантинного порядку та інші форс-мажорні обставини) в «Класичному приватному університеті».* Формат участі у дистанційних заняттях адаптується до поточних технічних можливостей та умов безпеки здобувачів і викладачів.

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

Основна література

1. Харченко В. О. Основи машинного навчання : навчальний посібник. Суми : Сумський державний університет, 2023. 264 с. URL: https://essuir.sumdu.edu.ua/bitstream-download/123456789/92711/3/Kharchenko_osnovy.pdf.

2. Кононова К. Ю. Машинне навчання: методи та моделі : підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка». Харків : ХНУ імені В. Н. Каразіна, 2020. 280 с. URL: https://karazinbook.com/sites/default/files/books/kononova_.pdf.

3. James G., Witten D., Hastie T., Tibshirani R., Taylor J. An Introduction to Statistical Learning: with Applications in Python**.** Cham : Springer, 2023. DOI: https://doi.org/10.1007/978-3-031-38747-0.

Додаткова література

4. Іванчук Я. В., Месюра В. І., Яровий А. А., Манжілевський О. Д. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних : навчальний посібник. Вінниця : ВНТУ, 2021. 69 с. URL: https://ir.lib.vntu.edu.ua/handle/123456789/36454.

5. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction**.** 2nd ed. New York : Springer, 2009. 745 p. DOI: https://doi.org/10.1007/978-0-387-84858-7.

6. Bishop C. M. Pattern Recognition and Machine Learning. New York : Springer, 2006. ISBN 978-0-387-31073-2. URL: https://link.springer.com/book/9780387310732.

7. Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction. 2nd ed. Cambridge, MA : The MIT Press, 2018. 552 p. URL: https://mitpress.mit.edu/9780262039246/reinforcement-learning/.

ІНФОРМАЦІЙНІ РЕСУРСИ В ІНТЕРНЕТІ

1. Scikit-learn. User Guide – документація з алгоритмів класифікації, регресії, кластеризації, підготовки даних та оцінювання моделей. URL: https://scikit-learn.org/stable/user_guide.html.

2. Scikit-learn. Common pitfalls and recommended practices – рекомендації щодо запобігання витоку даних, коректної підготовки вибірки та відтворюваності результатів. URL: https://scikit-learn.org/stable/common_pitfalls.html.

3. An Introduction to Statistical Learning – авторський сайт підручника та навчальні матеріали. URL: https://www.statlearning.com/.

4. ISL with Python: Resources – набори даних, презентації та матеріали до практичних занять. URL: https://www.statlearning.com/resources-python.

5. UCI Machine Learning Repository – репозитарій наборів даних для задач статистичного навчання. URL: https://archive.ics.uci.edu/.

6. Kaggle Datasets – платформа наборів даних для прикладних досліджень і навчальних проєктів. URL: https://www.kaggle.com/datasets.