



Міжнародний гуманітарний університет
Факультет кібербезпеки, програмної інженерії та комп'ютерних наук
Кафедра інформаційних технологій

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

МАШИННЕ НАВЧАННЯ

Галузь знань	<u>12 «Інформаційні технології»</u>
Спеціальність	<u>122 «Комп'ютерні науки»</u>
Назва освітньої програми	<u>Комп'ютерні науки</u>
Рівень вищої освіти	<u>другий (магістерський) рівень</u>

Розробники і викладачі	Контактний тел.	E-mail
викладач кафедри інформаційних технологій, доктор філософії PhD Радюк Павло Михайлович	097-854-91-48	radiukpavlo@gmail.com

1. АНОТАЦІЯ ДО КУРСУ

Курс призначений для підготовки магістрів з галузі знань 12 – Інформаційні технології за спеціальністю 122 – Комп'ютерні науки. На основі загальних понять вищої математики, ймовірнісних процесів та математичної статистики дисципліна розглядає основні властивості й операції над даними та множинами, методи та принципи машинного навчання, основні алгоритми та програмне забезпечення з метою розв'язання класу задач.

Метою викладання навчальної дисципліни є формування у студентів фундаментальних знань про принципи роботи з даними, про роботу з комбінаторними об'єктами, системами та моделями; оволодіння студентами математичним та інтелектуальним апаратами інженера програмного забезпечення.

2. ОЧІКУВАНІ КОМПЕТЕНТНОСТІ, ЯКІ ПЛАНУЄТЬСЯ СФОРМУВАТИ ТА ДОСЯГНЕННЯ ПРОГРАМНИХ РЕЗУЛЬТАТІВ

У процесі реалізації програми дисципліни «Машинне навчання» формуються наступні компетентності із передбачених освітньою програмою:

Інтегральна компетентність

ІК1. Здатність розв'язувати задачі дослідницького та/або інноваційного характеру у сфері комп'ютерних наук.

Загальні компетентності

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК05. Здатність вчитися й оволодівати сучасними знаннями.

ЗК07. Здатність генерувати нові ідеї (креативність).

Спеціальні (фахові) компетентності

СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК06. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.

СК07. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.

СК08. Здатність розробляти і реалізовувати проекти зі створення програмного забезпечення, у тому числі в непередбачуваних умовах, за нечітких вимог та необхідності застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проектом.

СК09. Здатність розробляти та адмініструвати бази даних та знань.

СК10. Здатність оцінювати та забезпечувати якість ІТ-проектів, інформаційних та комп'ютерних систем різного призначення, застосовувати міжнародні стандарти оцінки якості програмного забезпечення інформаційних та комп'ютерних систем, моделі оцінки зрілості процесів розробки інформаційних та комп'ютерних систем.

Навчальна дисципліна «Машинне навчання» забезпечує досягнення програмних результатів навчання, передбачених освітньою програмою:

РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

РН8. Розробляти математичні моделі та методи аналізу даних (включно з великим).

РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

РН10. Проектувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення

РН11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.

Заплановані результати навчання за навчальною дисципліною

Знання:

- принципи побудови векторів ознак, вирішальних правил і класифікації;
- основні види класифікаторів;
- принципи побудови лінійних та нелінійних класифікаторів;

- методи кластеризації;
- особливості вибору ознак класифікації і попередньої обробки даних;
- основні методи штучного інтелекту, засоби формалізованого подання знань, напрямки інтелектуалізації автоматизованих інформаційних та інформаційно-керуючих систем.

Уміння:

- вибирати відповідний вид класифікатора в залежності від розв’язуваної задачі;
- вибирати набір ознак для класифікації і проводити попередню обробку даних;
- застосовувати алгоритми побудови і навчання класифікатора по вибірці;
- виконувати обчислення, пов’язані з навчанням і роботою класифікатора.

Навички:

- розробляти алгоритми та програми для обробки баз даних та знань (пошук інформації, складання логічних рівнянь);
- працювати з інструментарієм аналітичної обробки даних, який є у складі корпоративними СКБД, сховищами даних та базами даних в Internet-середовищі;
- проводити чисельні експерименти на модельних і реальних даних та інтерпретувати їхні результати;
- представляти результати досліджень в усній і письмовій формах.

3. ОБСЯГ ТА ОЗНАКИ КУРСУ

Загалом		Вид заняття (денне відділення)			Ознаки курсу		
ЄКТС	годин	Лекційні заняття	Практичні заняття	Самостійна робота	Курс, (рік навчання)	Семестр	Обов’язкова / вибіркова
5	150	28/6	28/6	94/138	1	1	Обов’язкова

4. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин									
	Денна форма					Заочна форма				
	Усього	у тому числі				Усього	у тому числі			
		Лекц.	Прак.	Лаб.	Сам. роб.		Лекц.	Прак.	Лаб.	Сам. роб.
Тема 1. Первинний аналіз даних з Pandas.	10	2	2		6	10	2			8
Тема 2. Об’єкти та класи в Python.	10	2	2		6	10				10
Тема 3. Візуальний аналіз даних з Python.	10	2	2		6	10	2			8
Тема 4. Візуальна аналітика для задач машинного навчання.	10	2	2		6	10				10

Тема 5. Метод дерева рішень для задач класифікації та регресії.	10	2	2		6	10	2			8
Тема 6. Метод найближчих сусідів для задач класифікації та регресії.	10	2	2		6	10				10
Тема 7. Лінійні моделі регресії.	10	2	2		6	10				10
Тема 8. Лінійні моделі класифікації.	10	2	2		6	10				10
Усього за змістовний модуль 1	80	16	16		48	80				74
Тема 9. Ансамблі алгоритмів: беггінг.	10	2	2		6	10		2		8
Тема 10. Ансамблі алгоритмів: випадковий ліс.	12	2	2		8	12				12
Тема 11. Побудова та відбір ознак.	12	2	2		8	12		2		10
Тема 12. Основні метрики якості класифікації.	12	2	2		8	12				12
Тема 13. Методи пониження розмірності.	12	2	2		8	12		2		10
Тема 14. Методи кластерного аналізу.	12	2	2		8	12				12
Усього за змістовний модуль 2	70	12	12		46	70				64
Усього годин	150	28	28		94	150	6	6	-	138
ПІДСУМКОВИЙ КОНТРОЛЬ – ЕКЗАМЕН										

5. ТЕХНІЧНЕ Й ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ / ОБЛАДНАННЯ

Студенти отримують теми та питання курсу, основну і додаткову літературу, рекомендації, завдання та оцінки за їх виконання як традиційним шляхом, так і з використанням університетської платформи онлайн навчання на базі Moodle. Окрім того, практичні навички у пошуку та аналізу інформації за курсом, з оформлення індивідуальних завдань, тощо, студенти отримують, користуючись університетськими комп'ютерними класами та бібліотекою.

6. ПИТАННЯ ДО ПРАКТИЧНИХ ЗАНЯТЬ

№	Назва теми	Кількість годин	
		Денна форма	Заочна форма
1	Налаштування робочого середовища Python та Pandas.	2	
2	Аналіз даних за доходом населення UCI Adult.	2	
3	Налаштування бібліотек візуалізації даних середовища Python.	2	
4	Візуальний аналіз даних публікацій соціальних мереж.	2	
5	Функції для розрахунку ентропії та приросту інформації.	2	
6	Дерево рішень та метод найближчих сусідів на наборі даних UCI Adult.	2	
7	Побудова моделі класифікації на основі лінійної регресії.	2	

8	Прогнозування популярності статей на TechMedia за допомогою лінійних регресій.	2	
	Усього за змістовний модуль 1	16	-
9	Бегінг логістичних регресій для задачі кредитного скорингу.	2	2
10	Випадковий ліс для задачі кредитного скорингу.	2	
11	Лінійна регресія для задачі визначення якості вина.	2	2
12	Lasso-регресія та випадковий ліс для задачі визначення якості вина.	2	
13	Розв'язування задачі пониження розмірності.	2	2
14	Розв'язування задачі визначення видів активності людини.	2	
	Усього за змістовний модуль 2	12	
	Усього годин	28	6

7. САМОСТІЙНА РОБОТА

До самостійної роботи студентів щодо вивчення дисципліни «Машинне навчання» включаються:

1. Знайомство з науковою та навчальною літературою відповідно зазначених у програмі тем.
2. Опрацювання лекційного матеріалу.
3. Підготовка до практичних занять.
4. Консультації з викладачем протягом семестру.
5. Самостійне опрацювання окремих питань навчальної дисципліни.
6. Підготовка до підсумкового контролю.

Тематика та питання до самостійної підготовки та індивідуальних завдань

№	Назва теми	Кількість годин	
		Денна форма	Заочна форма
1	Тема 1. Первинний аналіз даних з Pandas. Розгляд проблеми опрацювання даних. Демонстрація основних методів Pandas. Читання з файлу і первинний аналіз. Індекссування та вилучення даних.	6	8
2	Тема 2. Групування даних. Зведені таблиці. Перетворення датафреймів. Розв'язування задачі прогнозування відтоку клієнтів.	6	10
3	Тема 3. Візуальний аналіз даних з Python. Розгляд проблеми інтерпретування аналізу даних. Візуалізація даних. Основні бібліотеки Python для візуального аналізу даних. Matplotlib. Seaborn. Plotly.	6	8
4	Тема 4. Візуальна аналітика для задач машинного навчання. Гістограми. Теплові мапи. Коробковий графік. Метод t-SNE.	6	10

5	Тема 5. Метод дерева рішень для задач класифікації та регресії. Розгляд проблеми отримання ціннісної інформації. Ентропія. Алгоритм побудови дерева рішень. Критерії якості розбиття для задачі класифікації. Дерево рішень з кількісними ознаками. випадки для дерев рішень. Бібліотека sklearn.	6	8
6	Тема 6. Метод найближчих сусідів для задач класифікації та регресії. Вибір параметрів моделі найближчих сусідів. Крос-валідація. Переваги та недоліки методу найближчих сусідів. Складні випадки для найближчих сусідів.	6	10
7	Тема 7. Лінійні моделі регресії. Розгляд задачі передбачення. Метод найменших квадратів. Метод максимальної правдоподібності. Розкладання помилки на зміщення і розкид. Регуляризація лінійної регресії. L2-регуляризація логістичної функції втрат.	6	10
8	Тема 8. Лінійні моделі класифікації. Формальна модель класифікації. Логістична регресія, як лінійний класифікатор. Логістична регресія з поліноміальними ознаками. Задача класифікації відгуків IMDB до фільмів. Криві валідації й навчання.	6	10
	<i>Усього за змістовний модуль 1</i>	48	74
9	Тема 9. Ансамблі алгоритмів: беггінг. Розгляд проблеми низького узагальнення моделі оброблення даних. Формалізація ансамблів. Бутстреп. Беггінг. Помилка out-of-bag. Перетворення ознак в багатовимірний простір. Зміщення.	6	8
10	Тема 10. Ансамблі алгоритмів: випадковий ліс. Алгоритм випадкового лісу. Порівняння випадкового лісу з деревом рішень і беггінгом. Параметри випадкового лісу. Варіація й декореляційний ефект. Важливість ознак на прикладі випадкового лісу.	8	12
11	Тема 11. Побудова та відбір ознак. Типи даних. Виявлення ознак. Перетворення ознак. Нормалізація та зміна розподілу. Заповнення пропусків. Статистичні підходи до відбору ознак.	8	10
12	Тема 12. Основні метрики якості класифікації. Матриця невідповідностей. Характеристики бінарного класифікатора. ROC-крива. Площа під кривою AUC. Дисбаланс класів. Приклади застосувань метрик класифікації. Багатокласова класифікація.	8	12
13	Тема 13. Методи пониження розмірності. Розгляд проблеми великої розмірності даних. Метод головних компонент. Приклад використання методу пониження розмірності. Метод t-SNE. Сингулярне розкладання матриці, SVD.	8	10
14	Тема 14. Методи кластерного аналізу. Розгляд проблеми оброблення неструктурованих даних. Метод k-середніх. Метод поширення спорідненості. Метод ієрархічної кластеризації. Вибір кількості кластерів. Метод ліктя. Метрики якості кластеризації.	8	12
	<i>Усього за змістовний модуль 2</i>	46	64
	<i>Усього годин</i>	94	138

8. ВИДИ ТА МЕТОДИ КОНТРОЛЮ

Робоча програма навчальної дисципліни передбачає наступні види та методи контролю:

Види контролю	Складові оцінювання
поточний контроль , який здійснюється у ході: проведення практичних занять, виконання варіантів завдань, проведення консультацій та відпрацювань, підсумковий контроль.	50%
підсумковий контроль , який здійснюється у ході проведення екзамену.	50%

Методи діагностики знань (контролю)	фронтальне опитування, індивідуальне опитування, робота у групах, розв'язання задач і практичних завдань, екзамен
--	---

8. ОЦІНЮВАННЯ ПОТОЧНОЇ, САМОСТІЙНОЇ ТА ІНДИВІДУАЛЬНОЇ РОБОТИ СТУДЕНТІВ З ПІДСУМКОВИМ КОНТРОЛЕМ У ФОРМІ ЕКЗАМЕНУ/ ЗАЛІКУ

Денна форма навчання / Заочна форма навчання			
<i>Поточний контроль</i>			
Види роботи	Планові терміни виконання	Форми контролю та звітності	Максимальний відсоток оцінювання
Систематичність і активність роботи на семінарських (практичних) заняттях			
1.1. Підготовка до практичних занять	Відповідно до робочої програми та розкладу занять	Перевірка обсягу та якості засвоєного матеріалу під час практичних занять	25
Виконання завдань для самостійного опрацювання			
1.2. Підготовка програмного матеріалу (тем, питань), що виноситься на самостійне вивчення	-//-	Розгляд відповідного матеріалу під час аудиторних занять або ІКР ¹ , перевірка конспектів навчальних текстів тощо	10
Виконання індивідуальних завдань (науково-дослідна робота студента)			
1.3. Підготовка реферату за заданою тематикою	Відповідно до розкладу занять і графіку ІКР	Обговорення (захист) матеріалів реферату	10
1.4. Інші види індивідуальних завдань, в т.ч. підготовка наукових публікацій, участь у роботі круглих столів, конференцій тощо.	-//-	Обговорення результатів проведеної роботи під час аудиторних занять або ІКР, наукових конференцій та круглих столів.	5
Разом балів за поточний контроль			50

¹ Індивідуально-консультаційна робота викладача зі студентами

Підсумковий контроль Екзамен	50
Усього балів	100

10. КРИТЕРІЇ ПІДСУМКОВОЇ ОЦІНКИ ЗНАНЬ СТУДЕНТІВ (для екзамену)

Рівень знань оцінюється:

- «відмінно»/ «зараховано» А - від 90 до 100 балів. Студент виявляє особливі творчі здібності, вміє самостійно знаходити та опрацьовувати необхідну інформацію, демонструє знання матеріалу, проводить узагальнення і висновки. Був присутній на лекціях та семінарських заняттях, під час яких давав вичерпні, обґрунтовані, теоретично і практично правильні відповіді, має конспект з виконаними завданнями до самостійної роботи, проявляє активність і творчість у науково-дослідній роботі;

- «добре»/ «зараховано» В - від 82 до 89 балів. Студент володіє знаннями матеріалу, але допускає незначні помилки у формуванні термінів, категорій, проте за допомогою викладача швидко орієнтується і знаходить правильні відповіді. Був присутній на лекціях та семінарських заняттях, має конспект з виконаними завданнями до самостійної роботи, проявляє активність і творчість у науково-дослідній роботі;

- «добре»/ «зараховано» С - від 74 до 81 балів. Студент відтворює значну частину теоретичного матеріалу, виявляє знання і розуміння основних положень, з допомогою викладача може аналізувати навчальний матеріал, але дає недостатньо обґрунтовані, невичерпні відповіді, допускає помилки. При цьому враховується наявність конспекту з виконаними завданнями до самостійної роботи та активність у науково-дослідній роботі;

- «задовільно»/ «зараховано» D - від 64 до 73 балів. Студент був присутній не на всіх лекціях та семінарських заняттях, володіє навчальним матеріалом на середньому рівні, допускає помилки, серед яких є значна кількість суттєвих. При цьому враховується наявність конспекту з виконаними завданнями до самостійної роботи;

- «задовільно»/ «зараховано» Е - від 60 до 63 балів. Студент був присутній не на всіх лекціях та семінарських заняттях, володіє навчальним матеріалом на рівні, вищому за початковий, значну частину його відтворює на репродуктивному рівні, на всі запитання дає необґрунтовані, невичерпні відповіді, допускає помилки, має неповний конспект з завданнями до самостійної роботи.

- «незадовільно з можливістю повторного складання» / «не зараховано» FX – від 35 до 59 балів. Студент володіє матеріалом на рівні окремих фрагментів, що становлять незначну частину навчального матеріалу.

- «незадовільно з обов'язковим повторним вивченням дисципліни» / «не зараховано» F – від 0 до 34 балів. Студент не володіє навчальним матеріалом.

Таблиця відповідності результатів контролю знань за різними шкалами

100-бальною шкалою	Шкала за ECTS	За національною шкалою	
		іспит	залік
90-100 (10-12)	A	Відмінно	зараховано
82-89 (8-9)	B	Добре	
74-81(6-7)	C		
64-73 (5)	D	Задовільно	
60-63 (4)	E		

35-59 (3)	FX	незадовільно	не зараховано
1-34 (2)	F		

11. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Могильний С.Б. Машинне навчання з використанням мікрокомп'ютерів : навч.-метод. посіб. / за ред. О. В. Лісового та ін. Київ, 2019. 224 с. URL: <https://api.man.gov.ua/api/assets/man/54c0ee59-b490-4ff3-a346-90a89fd67e30/>
2. Мосіюк О.О. Штучний інтелект: Вступ до машинного навчання : навч.-метод. посіб. / за рец. М.О. Медведєвої, І.С. Мінтія. Житомир: Вид-во ЖДУ ім. Івана Франка, 2019. 76 с. URL: <http://eprints.zu.edu.ua/32982/>
3. Кононова К.Ю. Машинне навчання: Методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка» / за рец. А.В. Матвійчука, Л.С. Гур'янової. Харків: ХНУ імені В. Н. Каразіна, 2020. 301 с. URL: https://www.researchgate.net/publication/345765254_MASINNE_NAVCANNA_METODI_TA_MODELI
4. Троцько В.В. Методи штучного інтелекту : навч.-методич. і практ. посіб. / за рец. А.М. Котенка, Ю.Є. Добришина. Київ: Університет економіки та права «КРОК», 2020 86 с. URL: https://library.krok.edu.ua/media/library/category/navchalni-posibniki/trotsko_0001.pdf
5. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посіб. / за рец. В.О. Філатова, О.А. Винокурової. Харків: ХНУРЕ, 2021. 92 с. URL: <https://openarchive.nure.ua/server/api/core/bitstreams/2e55d639-52fd-48d9-b7b7-14989f49f291/content>
6. Болюбаш Н.М. Методичні вказівки до виконання курсової роботи з дисципліни «Інтелектуальний аналіз» даних для студентів спеціальності 122 – Комп'ютерні науки : методичні вказівки / за рец. І.П. Атаманюка. Миколаїв : Вид-в ЧНУ ім. Петра Могили, 2021. 28 с. URL: <https://bit.ly/40sKIGw>

Допоміжна

7. Сергєєв-Горчинський О.О., Іщенко Г.В. Інтелектуальний аналіз даних: Комп'ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / за рец. В.С. Рогози. Київ : КПІ ім. Ігоря Сікорського, 2018. 73 с. URL: https://ela.kpi.ua/bitstream/123456789/24971/1/Komp_prakt.pdf
8. Tagliaferri L., Morales M., Birbeck E., Wan A. Python machine learning projects / ed. B. Hogan, M. Drake. Published under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0. New York City: DigitalOcean, 2019. 135 p. URL: <https://assets.digitalocean.com/books/python/machine-learning-projects-python.pdf>
9. Deisenroth M.P., Faisal A.A., Ong C. S. Mathematics for machine learning. Cambridge: Cambridge University Press, 2020. 412 p. URL: <https://mml-book.github.io/book/mml-book.pdf>
10. Єгорова О.В. Методичні рекомендації до самостійної роботи з дисципліни «Інтелектуальний аналіз даних» для здобувачів освітнього ступеня «бакалавр» зі спеціальностей 122 – Комп'ютерні науки та 126 – Інформаційні системи та технології усіх форм навчання [Електронний ресурс] / за рец. А.О. Лавданського. М-во освіти і науки України, Черкас. держ. технол. ун-т. Черкаси: ЧДТУ, 2020. 13 с. URL: <https://er.chdtu.edu.ua/bitstream/ChSTU/1605/1/19-5-25-%D0%9F.pdf>
11. Zhang A., Lipton Z.C., Li M., Smola A.J. Dive into deep learning: 3rd ed. arXiv preprint arXiv:2106.11342, 2021. 1197 p. URL:

<https://arxiv.org/abs/2106.11342>

12. Raschka S., Liu Y., Mirjalili V. Machine learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python. Birmingham, UK: Packt Publishing Ltd., 2022. 770 p. URL: <https://github.com/rasbt/machine-learning-book>

13. Banachewicz K., Massaron L. The Kaggle book: Data analysis and machine learning for competitive data science. Birmingham, UK: Packt Publishing Ltd., 2022. 531 p. URL: <https://github.com/PacktPublishing/The-Kaggle-Book>

14. Олещенко Л.М. Машинне навчання: Комп'ютерний практикум з дисципліни «Машинне навчання» [Електронний ресурс]: навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» / за рец. С.М. Чумаченка, А.О. Мошенського. Київ: КПІ ім. Ігоря Сікорського, 2022. 92 с. URL: https://ela.kpi.ua/bitstream/123456789/48804/1/Mashynne_navchannia.pdf

Інформаційні ресурси

15. Офіційний сайт Python [Електронний ресурс] URL: <https://www.python.org> (Дата звернення: 01.08.2023)

16. Schmidt, M. 80 lectures on machine learning. Mark Schmidt, University of British Columbia. 11.01.2016 Web. URL: <https://www.cs.ubc.ca/~schmidtm/Courses/LecturesOnML/> (Дата звернення: 01.08.2023)

17. Машинне навчання. Інженерія програмного забезпечення [Електронний ресурс] // Moodle. MGU. 2023. URL: <https://moodle.mgu.edu.ua/course/view.php?id=1199> (Дата звернення: 01.08.2023)