

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ І.І. МЕЧНИКОВА
Кафедра оптимального керування і економічної кібернетики

ЗАТВЕРДЖУЮ
Проректор з науково-педагогічної роботи
Майя НІКОЛАЄВА
«___» _____ 20__ р.



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

БК Методи Machine Learning

Рівень вищої освіти: Другий (магістерський)

Галузь знань: F Інформаційні технології

Спеціальність: F7 Комп'ютерна інженерія

Освітньо-професійна програма: Комп'ютерна інженерія

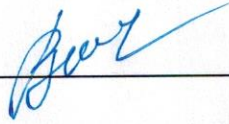
Робоча програма навчальної дисципліни «Методи Machine Learning». Одеса: ОНУ,
2026. – 18 с.

Розробник:

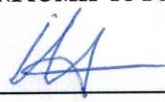
Страхов Євген Михайлович, кандидат фізико-математичних наук, доцент, доцент
кафедри оптимального керування і економічної кібернетики

Робоча програма затверджена на засіданні кафедри оптимального керування і
економічної кібернетики

Протокол № 1 від «31» серпня 2026 р.

Завідувач кафедри  Віктор ВЕРБИЦЬКИЙ

Погоджено із гарантом ОПП «Комп'ютерна інженерія»

 Юрій ГУНЧЕНКО

Схвалено Навчально-методичною комісією (НМК) з ІТ-спеціальностей факультету
математики, фізики та інформаційних технологій

Протокол № 1 від «28» вер 2026 р.

Голова НМК  Лариса МАРТИНОВИЧ

Переглянуто та затверджено на засіданні кафедри _____

Протокол № _____ від «____» _____ 20__ р.

Завідувач кафедри _____ (_____) (Ім'я ПРІЗВИЩЕ)
(підпис)

Переглянуто та затверджено на засіданні кафедри _____

Протокол № _____ від «____» _____ 20__ р.

Завідувач кафедри _____ (_____) (Ім'я ПРІЗВИЩЕ)
(підпис)

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, спеціальність, спеціалізація, рівень вищої освіти	Характеристика навчальної дисципліни	
		Очна форма навчання	Заочна форма навчання
Загальна кількість: кредитів – 3 годин – 90 змістових модулів – 3	Галузь знань F Інформаційні технології Спеціальність F7 Комп'ютерна інженерія ОПП/ОНП Комп'ютерна інженерія Рівень вищої освіти: <u>Другий (магістерський)</u>	<i>Вибіркова</i>	
		<i>Рік підготовки:</i>	
		2-й	-й
		<i>Семестр</i>	
		3-й	-й
		<i>Лекції</i>	
		16 год.	год.
		<i>Практичні, семінарські</i>	
		14 год.	год.
		<i>Лабораторні</i>	
		-	-
		<i>Самостійна робота</i>	
		60 год.	год.
		Форма підсумкового контролю: <i>залік</i>	

2. Мета та завдання навчальної дисципліни

Мета: формування у здобувачів поглиблених системних знань та практичних компетентностей у сфері розробки, оптимізації та впровадження складних моделей машинного навчання (Machine Learning); підготовка фахівців, здатних вирішувати нетривіальні дослідницькі та прикладні задачі Data Science, проводити порівняльний аналіз алгоритмів і проектувати ефективні ML-рішення.

Завдання:

- критичний аналіз та систематизація математичного апарату й алгоритмічних основ методів навчання із вчителем (supervised learning) та без вчителя (unsupervised learning);
- оволодіння методами поглибленого інжинірингу ознак (Feature Engineering), складного розвідувального аналізу (EDA) та обробки нестандартних чи зашумлених даних;
- проектування повного циклу розробки моделей (Machine Learning Pipeline): від вибору метрик оцінки якості й автоматизованого налаштування гіперпараметрів до оцінки стійкості, валідації та інтерпретації результатів;
- набуття навичок порівняльного аналізу й оптимізації швидкодії та точності моделей під специфічні бізнес- або наукові обмеження;
- удосконалення практичних навичок високорівневого програмування мовою Python із використанням сучасного стеку бібліотек (NumPy, SciPy, Pandas, Scikit-Learn, XGBoost/LightGBM, Optuna тощо) для розв'язання задач класифікації, регресії, кластеризації, зменшення розмірності та побудови рекомендаційних систем.

У результаті вивчення навчальної дисципліни здобувач вищої освіти повинен **знати:**

- Математичні та теоретичні основи сучасних методів машинного навчання (теорія ймовірностей, математична статистика, теорія оптимізації, лінійна алгебра в контексті ML);
- Алгоритмічні та практичні особливості складних ансамблевих методів (градієнтний бустинг, беггінг, стекінг), алгоритмів поглибленого зниження розмірності (PCA, t-SNE, UMAP) та кластеризації (HDBSCAN, GMM);
- Методології та патерни розробки наскрізних пайплайнів машинного навчання (End-to-End ML Pipelines);
- Методи поглибленого інжинірингу та відбору ознак (Feature Engineering & Feature Selection), а також підходи до обробки мультиколінеарності, незбалансованих та зашумлених даних;
- Принципи інтерпретованості моделей (Explainable AI / XAI: SHAP, LIME) та оцінки їхньої стійкості, робастності й узагальнюючої здатності;
- Методологію автоматизованої оптимізації гіперпараметрів та комплексної оцінки якості моделей за допомогою технічних та бізнес-метрик.

вміти:

- Проектувати та реалізовувати повний цикл розробки ML-моделей для розв'язання складних практичних задач класифікації, регресії, кластеризації, зниження розмірності та побудови рекомендаційних систем;
- Використовувати поглиблений стек інструментів Data Science мовою Python (NumPy, SciPy, Pandas, Scikit-Learn, XGBoost, LightGBM, Optuna, SHAP тощо) для обробки великих масивів даних;
- Проводити складний розвідувальний аналіз (Advanced EDA), автоматизовувати процес попередньої обробки даних та генерувати нові інформовані ознаки;
- Застосовувати алгоритми автоматичного пошуку та оптимізації гіперпараметрів моделі для досягнення максимальних показників якості;
- Здійснювати крос-валідацію та оцінку моделей із врахуванням специфіки розподілу даних (часові ряди, дисбаланс класів, ієрархічні дані);
- Аналізувати наукову літературу, документацію та репозиторії новітніх ML-бібліотек (зокрема англійською мовою) для інтеграції передових дослідницьких підходів у власні рішення.

бути здатним:

- Переводити реальні бізнес- або наукові проблеми у формальну математичну та алгоритмічну постановку задачі машинного навчання;
- Критично оцінювати та аргументовано обирати архітектуру моделі, алгоритми та метрики якості відповідно до поставленої задачі та наявних обчислювальних ресурсів (trade-off між точністю та швидкістю);
- Розробляти та тестувати наскрізні ML-пайплайни (End-to-End ML Solutions) — від первинного очищення та підготовки даних до отримання готової до інтеграції моделі;
- Пояснювати та інтерпретувати рішення складних ML-моделей («чорних скриньок») для забезпечення прозорості та довіри з боку замовника чи кінцевого користувача;
- Виявляти та усувати проблеми перенавчання (overfitting), недонавчання (underfitting), витоку даних (data leakage) та зміщення вибірки (data drift / concept drift);
- Проводити самостійне прикладне чи наукове дослідження, порівняльний бенчмаркінг різних підходів та формулювати обґрунтовані висновки.

Опанування цієї навчальної дисципліни забезпечує набуття здобувачами навичок і компетентностей, спрямованих на досягнення глобальних цілей сталого розвитку до 2030 року, що проголошені резолюцією Генеральної Асамблеї Організації Об'єднаних Націй від 25 вересня 2015 року No 70/1 та визначені Указом Президента України від 30 вересня 2019 року No 722, зокрема:

.....

Ціль 9: створення стійкої інфраструктури, сприяння всеохоплюючій і сталій індустріалізації та інноваціям.

3. Зміст навчальної дисципліни

Змістовий модуль 1. Поглиблений розвідувальний аналіз (Advanced EDA) та інжиніринг ознак (Feature Engineering)

Тема 1.1. Препроцесинг даних. Методології підготовки даних для промислових та дослідницьких задач. Стратегії обробки пропущених значень (Advanced Imputation). Виявлення та обробка аномалій/викидів (Outlier Detection: Isolation Forest, Local Outliers). Обробка часових та геопросторових даних. Проблема зміщення вибірки та витоку даних (Data Drift, Data Leakage).

Тема 1.2. Розширені методи візуалізації та багатовимірний статистичний аналіз. Візуалізація розподілів багатовимірних даних. Побудова та інтерпретація зашумлених категоріальних та числових ознак. Прийоми візуалізації даних. Кореляція проти каузальності (Correlation vs Causality). Нелінійні залежності та взаємна інформація (Mutual Information). Статистична перевірка гіпотез у ML. Відбір найбільш інформативних ознак (Feature Selection: Filter, Wrapper, Embedded methods).

Зв'язок із ЦСР 9 (Промисловість, інновації та інфраструктура):

Тематика модуля орієнтована на обробку часових рядів і телеметрії з датчиків промислового обладнання та об'єктів критичної інфраструктури (IoT). Це формує вміння оцінювати якість даних для моніторингу надійності інфраструктурних систем.

Змістовий модуль 2. Навчання із вчителем (Supervised Learning), ансамблі та інтерпретованість моделей

Тема 2.1. Математичні основи, мінімізація ризику та ML-пайплайни. Формалізація задач класифікації та регресії. Теоретичні основи оптимізації та функцій втрат (Loss & Cost functions). Проблема перенавчання (Overfitting), компроміс зсуву та дисперсії (Bias-Variance Tradeoff). Стратегії валідації складних даних (Stratified, GroupKFold, TimeSeriesSplit). Проектування наскрізних пайплайнів (Scikit-Learn Pipelines).

Тема 2.2. Метричні алгоритми та структури просторового індексування. Оптимізація обчислень у метричних методах (k-d trees, Ball Tree). Метод kNN у задачах великої розмірності. Вплив нормалізації даних та метрик відстані на стійкість рішень.

Тема 2.3. Деревя рішень, ансамблеві методи та градієнтний бустинг. Деревя рішень: критерії розщеплення (Gini, Entropy), усічення (Pruning). Поліпшення

стійкості: Random Forest, Bagging. Сучасні бібліотеки градієнтного бустингу: XGBoost, LightGBM, CatBoost. Стекінг та блендинг моделей (Stacking & Blending).

Тема 2.4. Лінійні моделі та найпростіші нейронні мережі. Лінійна та логістична регресія. Проблема мультиколінеарності. Штрафна регуляризація (Ridge, Lasso, ElasticNet). Машини опорних векторів (SVM), ядерний трюк (Kernel Trick) та його застосування до нелінійних даних. Спеціалізовані метрики якості (ROC-AUC, PR-AUC, F-beta, custom loss). Автоматизований пошук гіперпараметрів (Grid/Random Search, Bayesian Optimization via Optuna). Нейронні мережі: від перцептрона до багатошарових мереж. Градієнтний спуск (SGD, Adam) та алгоритм зворотного поширення помилки (Backpropagation).

Тема 2.5. Інтерпретованість моделей. Проблема «чорної скриньки» в ML. Методи пояснення глобальної та локальної поведінки моделей: SHAP (Shapley Additive exPlanations), LIME, Partial Dependence Plots (PDP). Оцінка важливості ознак (Feature Importance).

Змістовий модуль 3. Навчання без вчителя (Unsupervised Learning) та рекомендаційні системи

Тема 3.1. Сучасні алгоритми кластеризації та виявлення аномалій. Ієрархічні, центроїдні та щільнісні методи кластеризації (K-Means++, DBSCAN, HDBSCAN). Ймовірнісні моделі (Gaussian Mixture Models — GMM). Оцінка якості кластеризації за відсутності розмітки (Silhouette score, Davies-Bouldin index).

Тема 3.2. Навчання представлень та зниження розмірності. Лінійні та нелінійні методи зниження розмірності: PCA, t-SNE, UMAP. Використання методів зниження розмірності для стиснення даних, видалення шуму та підготовки ознак.

Тема 3.3. Прикладні рекомендаційні системи. Collaborative & Content-Based Filtering. Матричні факторизації (SVD, ALS). Робота з неявним відгуком (Implicit Feedback). Метрики оцінки якості рекомендацій (NDCG, MAP@K, MRR).

Зв'язок із ЦСР 9 (Промисловість, інновації та інфраструктура):

Кластеризація та виявлення аномалій використовуються для розробки стійких цифрових систем, виявлення збоїв у мережевій інфраструктурі, класифікації режимів роботи технологічних ліній та оптимізації логістичних ланцюгів на підприємствах.

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин									
	Очна форма					Заочна форма				
	Усь ого	у тому числі				Усь ого	у тому числі			
		л	п/с	лаб	ср		л	п/с	лаб	ср
1	2	3	4	5	6	7	8	9	10	11
Змістовий модуль 1. Поглиблений розвідувальний аналіз (Advanced EDA) та інжиніринг ознак (Feature Engineering)										
Тема 1.1 Препроцесинг даних	9	1	2		6					
Тема 1.2. Розширені методи візуалізації та багатовимірний статистичний аналіз	9	1	2		6					
Разом за змістовим модулем 1	18	2	4		12					
Змістовий модуль 2. Навчання із вчителем (Supervised Learning), ансамблі та інтерпретованість моделей										
Тема 2.1. Математичні основи, мінімізація ризику та ML-пайплайни	8	2			6					
Тема 2.2. Метричні алгоритми та структури просторового індексування	10	2	2		6					
Тема 2.3. Дерева рішень, ансамблеві методи та градієнтний бустинг	10	2	2		6					
Тема 2.4. Лінійні моделі та найпростіші нейронні мережі	10	2	2		6					
Тема 2.5. Інтерпретованість моделей	6				6					
Разом за змістовим модулем 2	44	8	6		30					
Змістовий модуль 3. Навчання без вчителя (Unsupervised Learning) та рекомендаційні системи										
Тема 2.1. Сучасні алгоритми кластеризації та виявлення аномалій	10	2	2		6					
Тема 2.2. Навчання представлень та зниження розмірності	9	2	1		6					
Тема 2.3. Прикладні рекомендаційні системи	9	2	1		6					
Разом за змістовим модулем 3	28	6	4		18					
Усього годин	90	16	14		60					

5. Теми семінарських занять

Семінарські заняття не передбачені.

6. Теми практичних занять

№ з/п	Назва теми/види занять	Кількість годин	
		Очна форма	Заочна форма
1	Препроцесинг даних	2	
2	Розширені методи візуалізації та багатовимірний статистичний аналіз	2	
3	Метричні алгоритми	2	
4	Дерева рішень, ансамблеві методи та градієнтний бустинг	2	
5	Лінійні моделі та найпростіші нейронні мережі	2	
6	Сучасні алгоритми кластеризації та виявлення аномалій	2	
7	Навчання представлень та зниження розмірності. Рекомендаційні системи	2	
Усього годин		14	

7. Теми лабораторних занять

Лабораторні заняття не передбачені.

8. Самостійна робота

№ з/п	Назва теми/види занять	Кількість годин	
		Очна форма	Заочна форма
1	Препроцесинг даних. Стратегії обробки пропущених значень (Advanced Imputation). Виявлення та обробка аномалій/викидів (Outlier Detection: Isolation Forest, Local Outliers). Обробка часових та геопросторових даних. Проблема зміщення вибірки та витoku даних (Data Drift, Data Leakage). [1][3]	6	
2	Візуалізація даних та багатовимірний статистичний аналіз. Прийоми візуалізації даних. Кореляція проти каузальності (Correlation vs Causality). Нелінійні залежності та взаємна інформація (Mutual Information). Статистична перевірка гіпотез у ML. Відбір найбільш інформативних ознак (Feature Selection: Filter, Wrapper, Embedded methods). [1][2][3]	6	
3	Теоретичні основи машинного навчання. Формалізація задач класифікації та регресії. Теоретичні основи оптимізації та функцій втрат (Loss & Cost functions). Проблема перенавчання (Overfitting), компроміс зсуву та дисперсії (Bias-Variance Tradeoff). Стратегії валідації складних даних (Stratified, GroupKFold, TimeSeriesSplit). [1][3]	6	
4	Метричні алгоритми. Метод kNN у задачах великої розмірності. Вплив нормалізації даних та метрик відстані на стійкість рішень. [1][2][3]	6	
5	Дерева рішень, ансамблеві методи та градієнтний бустинг. Практична реалізація. [1][2][3]	6	
6	Лінійні моделі та найпростіші нейронні мережі. Машини опорних векторів (SVM), ядерний трюк (Kernel Trick) та його застосування до нелінійних даних. Автоматизований пошук	6	

	гіперпараметрів (Grid/Random Search, Bayesian Optimization via Optuna). [1][2][3]		
7	Інтерпретованість моделей. Проблема «чорної скриньки» в ML. Методи пояснення глобальної та локальної поведінки моделей: SHAP (Shapley Additive exPlanations), LIME, Partial Dependence Plots (PDP). Оцінка важливості ознак (Feature Importance). [1][3]	6	
8	Сучасні алгоритми кластеризації та виявлення аномалій. Ймовірнісні моделі (Gaussian Mixture Models — GMM). Оцінка якості кластеризації за відсутності розмітки (Silhouette score, Davies-Bouldin index). [1][2][3]	6	
9	Навчання представлень та зниження розмірності. Лінійні та нелінійні методи зниження розмірності: PCA, t-SNE, UMAP. Використання методів зниження розмірності для стиснення даних, видалення шуму та підготовки ознак. [1][2][3]	6	
10	Рекомендаційні системи. Матричні факторизації (SVD, ALS). Робота з неявним відгуком (Implicit Feedback). Метрики оцінки якості рекомендацій (NDCG, MAP@K, MRR). [1][3]	6	
Усього годин		60	

До самостійної роботи відноситься:

- [1] – підготовка до лекцій, практичних, семінарських, лабораторних занять;
- [2] – індивідуальні та колективні проєктні завдання;
- [3] – неформальна та/або інформальна освіта тощо

Визнання результатів, отриманих в системі неформальної/інформальної освіти

В рамках опанування цієї навчальної дисципліни здобувач може пройти освітні або професійні курси/тренінги, професійне стажування, онлайн-курси на масових платформах чи у системі громадянської освіти.

Орієнтовний перелік курсів, які дотичні змісту цієї навчальної дисципліни:

Назва освітньої платформи	Назва курсу	Посилання
DataCamp	Machine Learning Fundamentals in Python	https://app.datacamp.com/learn/skill-tracks/machine-learning-fundamentals-with-python
DataCamp	Supervised Machine Learning in Python	https://app.datacamp.com/learn/skill-tracks/supervised-machine-learning-in-python
DataCamp	Machine Learning in Production in Python	https://app.datacamp.com/learn/skill-tracks/machine-learning-in-production
DataCamp	Data Manipulation in Python	https://app.datacamp.com/learn/skill-tracks/data-manipulation-with-python
DataCamp	Time Series in Python	https://app.datacamp.com/learn/skill-tracks/time-series-with-python
DataCamp	Importing & Cleaning Data in Python	https://app.datacamp.com/learn/skill-tracks/importing-cleaning-data-with-python

DataCamp	Applied Statistics in Python	https://app.datacamp.com/learn/skill-tracks/applied-statistics
----------	------------------------------	---

Здобувач має право також обрати самостійно освітню платформу та курс, які відповідають змісту цієї навчальної дисципліни.

Результати, отримані в системі неформальної/інформальної освіти, засвідчуються відповідним сертифікатом, де вказуються обсяг годин (або кількість кредитів ЄКТС) та здобуті компетентності та результати навчання. Такі результати можуть бути зараховані як результати поточного, періодичного або підсумкового контролю за цією навчальною дисципліною за процедурами, визначеними «Положенням про порядок визнання у вищій та фаховій передвищій освіті результатів навчання, здобутих шляхом неформальної та/або інформальної освіти в Одеському національному університеті імені І. І. Мечникова та відокремленому структурному підрозділі «Фаховий коледж ОНУ імені І. І. Мечникова» https://onu.edu.ua/pub/bank/userfiles/files/documents/polozennya/polozhennya_neformalna_osvita_2024_final.pdf

9. Методи навчання

У процесі вивчення навчальної дисципліни використовуються наступні методи навчання:

- Метод проблемного викладення (наукового пошуку).
- Пояснювально-ілюстративні методи: лекція, пояснення, самостійне опрацювання літературних джерел, робота з електронними конспектами лекцій та презентаціями, опрацювання наукових публікацій.
- Наочні методи (презентації, ілюстрації).
- Репродуктивні методи: розв'язування задач за алгоритмами конкретних методів, практичні роботи.
- Дослідницький метод.
- Методи формування і стимулювання пізнавальної діяльності: навчальні дискусії.

10. Форми контролю і методи оцінювання (у т.ч. критерії оцінювання результатів навчання)

Поточний контроль: індивідуальне усне поточне опитування

Періодичний контроль: усний захист практичних робіт, поточне тестування

Підсумковий контроль – залік (підсумкова оцінка виставляється за накопичувальною системою)

Критерії оцінювання за різними видами роботи

Вид контролю	Кількість балів
Опитування на лекціях та практичних заняттях	Від 0 до 2 балів
Усний захист практичної роботи	Від 0 до 8 балів
Поточні контрольні роботи (тестування)	Від 0 до 10 балів

Вид контролю	Бали	Критерії
Опитування на лекціях та практичних заняттях	0	Студент не орієнтується у навчальному матеріалі, робить грубі помилки або не може відповісти на прості питання
	1	Студент в цілому орієнтується в навчальному матеріалі, але при відповідях присутні неточності та/або логічні помилки
	2	Студент добре розуміє навчальний матеріал, може грамотно сформулювати та пояснити свою відповідь
Усний захист практичної роботи	0-4	Робота виконана не повністю, має суттєві недоліки і потребує допрацювання
	5	Робота виконана у повному обсязі, здобувач не впевнено орієнтується в коді, не може обґрунтувати свої дії та висновки, має суттєві недоліки або виявляє нерозуміння при відповідях на теоретичні питання
	6	здобувач впевнено орієнтується в коді, проте не завжди може обґрунтувати свої дії та висновки, має незначні недоліки або при відповідях на теоретичні питання
	7	Робота виконана у повному обсязі, здобувач в цілому орієнтується в коді та розуміє зміст виконаних дій, може обґрунтувати свої дії та висновки, має недоліки при відповідях на теоретичні питання
	8	Робота виконана у повному обсязі, здобувач розуміє суть моделі, відповідає на всі теоретичні питання, орієнтується в коді та розуміє зміст виконаних дій, може обґрунтувати свої дії та висновки
Поточні контрольні роботи (тестування)	0-10	Кількість балів дорівнює кількості правильних відповідей на тестові питання

Політика щодо дедлайнів та перескладання: У випадку пропуску заняття з поважних причин здобувач зобов'язаний виконати практичну роботу за відповідною темою у позанавчальний час за погодженням з викладачем. Перескладання відбувається із дозволу деканату за наявності поважних причин.

Політика щодо академічної доброчесності: під час опанування навчальної дисципліни категорично не припускаються будь-які види порушень академічної доброчесності з боку як здобувачів вищої освіти, так й викладача. У разі виявлення та фіксації будь-якого виду порушення академічної доброчесності, до порушника будуть застосовані санкції, передбачені Законом України «Про академічну доброчесність» та внутрішніми нормативними документами Одеського національного університету імені І. І. Мечникова.

Політика щодо використання штучного інтелекту: в ході опанування цієї навчальної дисципліни забороняється оприлюднення текстів, зображень, моделей, даних, інших результатів, згенерованих штучним інтелектом, як результатів власної

академічної та навчальної діяльності, якщо це не дозволено умовами навчального завдання, а сам цей факт не зазначено в навчальному завданні чи супровідних матеріалах до нього. Під час проходження підсумкового оцінювання використання ресурсів штучного інтелекту забороняється.

11. Питання до підсумкового контролю

Розвідувальний аналіз даних.

1. Попередня обробка даних. Описові статистики.
2. Типи даних. Приклади числових, рангових та категоріальних ознак.
3. Робота з пропущеними значеннями. Виявлення помилок.
4. Основні принципи візуалізації одновимірних даних. Типи діаграм для кількісних та категоріальних даних.
5. Основні принципи візуалізації двовимірних даних.
6. Візуалізація категоріальних даних.
7. Візуалізація багатовимірних даних на площині, прийоми «підвищення розмірності» графіка.
8. Кореляційний аналіз кількісних даних.
9. Кореляційний аналіз категоріальних даних.
10. Перевірка гіпотез про статистичну значущість взаємозв'язків.

Навчання із вчителем (supervised learning).

11. Математична та інтуїтивна постановки задачі навчання із вчителем.
12. Класифікація та регресія. Приклади.
13. Основні поняття та терміни машинного навчання.
14. Функція втрат. Функціонал якості. Приклади.
15. Параметри та гіперпараметри.
16. Проблема перенавчання.
17. Крос-валідація.
18. Метричні методи машинного навчання. Поняття відстані між об'єктами у просторі ознак.
19. Метод k найближчих сусідів (kNN).
20. Метод RadiusNeighbors.
21. Підбір оптимальної кількості сусідів.
22. Переваги та недоліки метричних методів.
23. Дерева рішень. Жадібний алгоритм побудови дерева рішень.
24. Поняття інформації та ентропії.
25. Проблеми перенавчання дерев.
26. Способи «обрізки» дерев.
27. Метод бутстреп. Ансамблі дерев, random forest.
28. Переваги та недоліки моделей на основі дерев рішень.
29. Лінійна регресія.
30. Проблема мультиколінеарності. Регуляризація лінійної регресії, ridge та lasso-регресія.
31. Логістична регресія. Регуляризація логістичної регресії.
32. Машини опорних векторів (SVM).

33. Підбір гіперпараметрів лінійних моделей.
 34. Переваги та недоліки лінійних моделей.
 35. Історія виникнення нейронних мереж.
 36. Функції активації.
 37. Лінійна регресія як тонка нейронна мережа.
 38. Логістична регресія як тонка нейронна мережа.
 39. Глибокі мережі.
 40. Алгоритм навчання мереж. Forward propagation, backpropagation.
 41. Основні метрики для задачі регресії та їх практичні аспекти.
 42. Метрики для задач класифікації. Матриця помилок.
 43. ROC-крива. Метрики на основі площі під кривою.
 44. Ансамблі моделей. Типи ансамблів у машинному навчанні.
 45. Бегінг.
 46. Бустинг. Градієнтний бустинг.
 47. Стекінг моделей.
- Навчання без вчителя (unsupervised learning).**
48. Методи кластеризації. Постановка задачі кластеризації. Критерії якості кластеризації.
 49. Ієрархічні методи, їх переваги та недоліки.
 50. Ітераційні методи. Метод k-середніх та його модифікації.
 51. Суть задачі зниження виміру та її прикладне значення.
 52. Метод головних компонент.
 53. Метод t-SNE.
 54. Багатовимірне метричне шкалювання.
 55. Рекомендаційні системи. Постановка задачі.
 56. Природа вхідних даних для задач побудови рекомендацій.
 57. Наївні методи побудови рекомендацій.
 58. Факторизація матриць. Рекомендаційні системи на основі машинного навчання.

12. Розподіл балів, які отримують здобувачі вищої освіти

Поточний та періодичний контроль										Сума балів
Змістовий модуль 1		Змістовий модуль 2					Змістовий модуль 3			
T1.1	T1.2	T2.1	T2.2	T2.3	T2.4	T2.5	T3.1	T3.2	T3.3	100
10	10	10		10	10		10		10	
КР за ЗМ 1 – 10		КР за ЗМ 2 – 10					КР за ЗМ 3 – 10			

*Примітка: Контрольна робота (КР) за змістовим модулем (ЗМ) здійснюється у формі письмових тестових завдань після завершення вивчення навчального матеріалу кожного змістового модуля. Тестові письмові завдання складаються з 10 тестових завдань і відповідають змісту навчального матеріалу модуля. За кожну правильну відповідь на одне тестове завдання студент отримує 1 бал

Шкала оцінювання: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90-100	A	відмінно	зараховано
85-89	B	добре	
75-84	C		
70-74	D	задовільно	
60-69	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

Оцінка за національною шкалою	Теоретична підготовка	Практична підготовка
	Здобувач освіти	
90-100/ зараховано	У повному обсязі та системно володіє теоретичним матеріалом. Вільно орієнтується в математичних та алгоритмічних основах ML (опукла оптимізація, функції втрат, теорія ансамблювання, ХАІ). Під час усних відповідей та тестування демонструє глибоке розуміння причинно-наслідкових зв'язків, аргументовано порівнює ML-моделі, здатен обґрунтувати вибір алгоритмів для складних прикладних задач та інфраструктурних проектів (зокрема в контексті Цілі сталого розвитку 9).	Глибоко та всебічно розкриває сутність практичних завдань. Самостійно та безпомилково розробляє наскрізні ML-пайплайни (End-to-End) мовою Python (Scikit-Learn, XGBoost/LightGBM, Optuna, SHAP). Демонструє високий рівень інжинірингу ознак (Feature Engineering), здійснює грамотну крос-валідацію, оцінку якості за технічними та бізнес-метриками. Виявляє творчий підхід, ефективно запобігає перенавчанню (overfitting) та витoku даних (data leakage).
75-89 / зараховано	Достатньо повно володіє навчальним матеріалом. Добре орієнтується в класичних та сучасних методах машинного навчання, володіє понятійним апаратом. Успішно складає модульне тестування та відповідає на усні запитання. При висвітленні складних теоретичних питань (наприклад, деталей опуклої оптимізації чи виведення математичних критеріїв розщеплення) може припускатися окремих несуттєвих неточностей, які самостійно виправляє під час уточнень.	Правильно вирішив більшість практичних завдань. Володіє стійкими навичками розробки ML-моделей та обробки даних у Python. Вміє будувати та налаштовувати класичні моделі й ансамблі, проводити базовий та розширений EDA. Під час виконання найбільш складних завдань (наприклад, побудови стекінг-ансамблів або налаштування байєсівської оптимізації) може потребувати незначних підказок викладача, але спроможний самостійно виправити помилки.
60-74 /зараховано	Володіє навчальним матеріалом на репродуктивному рівні. Знає базові концепції,	Може використовувати знання в стандартних ситуаціях. Має

	типи задач ML (класифікація, регресія, кластеризація) та стандартні метрики оцінки якості. При відповідях на усних опитуваннях та у тестуванні демонструє знання основних понять, але має суттєві труднощі при поясненні математичних принципів роботи складних моделей (бустинг, нейронні мережі) чи методів інтерпретованості. Допускає окремі істотні помилки.	елементарні, нестійкі навички виконання практичних робіт. Здатен за готовим шаблоном або алгоритмом навчити базову модель у Scikit-Learn та порахувати метрики. Має ускладнення при самостійному проведенні інжинірингу ознак, підборі гіперпараметрів, виправленні помилок у коді та обґрунтуванні вибору моделі.
35-59 / не зараховано з можливістю повторного складання	Володіє навчальним матеріалом поверхнево й фрагментарно. Під час усних опитувань та тестування допускає суттєві помилки, плутає фундаментальні поняття (наприклад, перенавчання та недонавчання, вибірки train/test/validation). Не здатний пояснити алгоритм роботи більшості моделей та обґрунтувати застосування метрик.	Недостатньо розкриває сутність практичних завдань. Припускається системних помилок у коді Python та логіці побудови ML-моделей. Не вміє самостійно підготувати датасет до навчання або обрати відповідний алгоритм. Здатний виконати лише окремі найпростіші операції за постійної підказки викладача; відсутні сформовані практичні навички.
0-34/ не зараховано з обов'язковим повторним вивченням дисципліни	Не володіє навчальним матеріалом. Демонструє повне незнання теоретичного базису дисципліни під час усних опитувань та тестування, не орієнтується у базовій термінології Data Science та Machine Learning.	Виконує лише окремі елементи завдань або повністю не виконує практичні роботи. Не володіє базовим стеком бібліотек Python для ML, не здатний запустити чи написати найпростіший скрипт для навчання моделі навіть за наявності детальних інструкцій.

Примітка: максимальна кількість балів за кожною темою вказана в п.12.

13.Навчально-методичне забезпечення

1. Робоча програма дисципліни.
2. Інтерактивні блокноти Jupyter notebook із текстом лекцій та програмним кодом.
3. Тестові завдання для періодичного контролю.

14.Рекомендована література

Основна

1. Mitchell, T. M. (1997). Machine Learning. McGraw-Hill, New York.
2. Murphy, K. P. (2012). Machine Learning: a Probabilistic Perspective. MIT Press, Cambridge, MA, USA.
3. Stephen Marsland. Machine Learning: An Algorithmic Perspective, 452 p., 2015.
4. Christopher M Bishop. Pattern recognition. Machine Learning, 128 p., 2006.
5. Ethem Alpaydin. Introduction To Machine Learning, 584 p., 2009.
6. Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. MIT Press.
7. Yaser S. Abu-Mostafa. Learning from data, 215 p., 2017
8. Alex Smola. Introduction to Machine Learning, 234 p., 2008.
9. Hastie, T., Tibshirani, R., and Friedman, J. (2001). The elements of statistical

- learning: data mining, inference and prediction. Springer Series in Statistics. Springer Verlag.
10. Hand D., Manila H., and Smyth P. (2001) Principles of Data Mining, MIT Press.
 11. Кононова К. Ю. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка» / К. Ю. Кононова. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с.
 12. Машинне навчання: комп'ютерний практикум з дисципліни «Машинне навчання» [Електронний ресурс]: навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем») / Л.М. Олещенко; КПІ ім. Ігоря Сікорського. – Електронні текстові дані. – Київ: КПІ ім. Ігоря Сікорського, 2022. – 92 с.

Додаткова

1. Shanyhin Anton, Babienko Volodymyr, Strakhov Yevhen, Korkhova Arina. Mathematical modeling of the dependence of the risk of vitamin D deficiency on anthropometric and laboratory parameters. Journal of Education, Health and Sport [online]. 28 April 2023, T. 13, nr 4, s. 356–366. [accessed 21.5.2023]. DOI 10.12775/JEHS.2023.13.04.042.
2. V. V. Babienko, A. V. Shanyhin, A. M. Rozhnova, Ye. M. Strakhov, A. S. Korkhova. Dependence of Vitamin D Level on Laboratory and Anthropometric Indicators: Application of Machine Learning Methods for Screening in Adults. Odesa medical journal. № 5 (190) 2024. P. 74–78. (Scopus)
<https://journals.onmedu.od.ua/index.php/med/article/view/160>
<https://doi.org/10.32782/2226-2008-2024-5-12>
3. Ryabov, M.I., Sukharev, A.L., Orlyuk, M.I., Ryabov, D.M., Sumaruk, Yu.P., Romenets, A.O., Strakhov, E.M., & Zabora, D. (2024). Space weather effects in Odesa magnetic anomaly — analysis by neural networks and wavelet methods. Astronomical and Astrophysical Transactions, 34(4), 299—316.
<https://doi.org/10.17184/eac.9276>

15. Електронні інформаційні ресурси

1. <http://lib.onu.edu.ua/> – Сайт бібліотеки ОНУ імені І. І. Мечникова;
2. <http://odnb.odessa.ua/> – Сайт Одеської національної наукової бібліотеки;
3. <https://cocalc.com/> — інтерактивне середовище для наукових обчислень та оформлення документації за допомогою популярних пакетів та мов програмування (Python, R, SageMath, Octave, LaTeX)
4. <https://colab.research.google.com/> — інтерактивне середовище для роботи з Jupyter notebook
5. <https://www.anaconda.com/> — дистрибутив математичних пакетів Python/R, а також пакетів для аналізу та візуалізації даних
6. <https://www.kaggle.com/> – ресурс для вивчення Data Science, пошуку датасетів, участі у змаганнях з машинного навчання, інтерактивне середовище для роботи з Jupyter notebook

7. <https://numpy.org/> – документація бібліотеки NumPy
8. <https://scipy.org/> – документація бібліотеки SciPy
9. <https://matplotlib.org/> – документація бібліотеки Matplotlib
10. <https://seaborn.pydata.org/> – документація бібліотеки Seaborn
11. <https://scikit-learn.org/> – документація бібліотеки Scikit-Learn
12. <https://www.deeplearningbook.org/> – підручник з глибинного навчання