



## Силабус освітнього компонента

Програма навчальної дисципліни



# Machine learning

**Шифр та назва спеціальності**

122 – Комп'ютерні науки

**Інститут**

ІНІ Комп'ютерного моделювання, прикладної фізики та математики

**Освітня програма**

Комп'ютерні науки. Моделювання, проектування та комп'ютерна графіка

**Кафедра**

Комп'ютерне моделювання процесів та систем (162)

**Рівень освіти**

Бакалавр

**Тип дисципліни**

Профільований пакет 4, Вибіркова

**Семестр**

6

**Мова викладання**

Українська

## Викладачі, розробники



**Багмут Іван Олександрович**

[ivan.bagmut@khp.edu.ua](mailto:ivan.bagmut@khp.edu.ua)

кандидат технічних наук, доцент

Спеціаліст у галузі науки про дані, машинного навчання та штучного інтелекту. Автор понад 40 наукових статей і матеріалів доповідей, співавтор методичних посібників.

[Детальніше про викладача на сайті кафедри](#)

## Загальна інформація

### Анотація

Навчальна дисципліна призначена для ознайомлення студентів з ключовими принципами та методами машинного навчання, включаючи навчання з вчителем, без вчителя та з підкріпленням. Велика увага приділяється практичному застосуванню теоретичних засад на реальних датасетах, а також розвитку вмінь вибору, налаштування та оцінки ефективності моделей машинного навчання, з метою підготовки студентів до вирішення складних задач у різноманітних галузях за допомогою машинного навчання.

### Мета та цілі дисципліни

Надання студентам знань та практичних навичок у розробці, аналізі та застосуванні моделей машинного навчання для вирішення реальних задач обробки даних.

### Формат занять

Лекції, лабораторні роботи, самостійна робота, консультації. Підсумковий контроль – екзамен.

### Компетентності

ЗК2: Здатність застосовувати знання у практичних ситуаціях.

ЗК3: Знання та розуміння предметної області та розуміння професійної діяльності.

СК11: 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.

СК19: Здатність застосовувати сучасні математичні концепції та алгоритмічні стратегії у сфері штучного інтелекту та машинного навчання для розробки новітніх моделей та систем, які здатні ефективно аналізувати, інтерпретувати, обробляти та використовувати складні дані, орієнтуючись на розширення та вдосконалення існуючих методів та технологій штучного інтелекту

## Результати навчання

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.

ПР20: Застосовувати вдосконалені математичні та алгоритмічні знання в області штучного інтелекту для створення інноваційних моделей та систем, які спроможні комплексно аналізувати та інтерпретувати складні та багатовимірні дані, відкриваючи нові можливості для поліпшення та оптимізації інтелектуальних технологій

## Обсяг дисципліни

Загальний обсяг дисципліни 120 год. (4 кредити ECTS): лекції – 32 год., лабораторні роботи – 16 год., самостійна робота – 72 год.

## Передумови вивчення дисципліни (пререквізити)

Для успішного освоєння курсу необхідні знання набути в дисциплінах "Лінійна алгебра", "Спеціальні глави вищої математики", "Мультипарадигмальні мови програмування", "Математичні методи теорії штучного інтелекту".

## Особливості дисципліни, методи та технології навчання

Основною мовою програмування, що використовується в рамках курсу, є мова Python. В якості середовища розробки обрано безкоштовне хмарне середовище Google Colab (Colaboratory), що дозволяє писати код в Jupyter notebooks.

Лекції ведуться з активним використанням мультимедійних ресурсів та інтерактивних методів навчання, що включає аналіз прикладів, кейсів та реальних проектів. Матеріал подається через демонстраційний підхід з акцентом на систематизацію ключових концепцій, а також залучення студентів до дискусій та критичного аналізу.

Навчальні матеріали доступні студентам на Microsoft OneDrive.

## Програма навчальної дисципліни

### Теми лекційних занять

#### Тема 1. Вступ до машинного навчання

Огляд основних понять та типів машинного навчання. Основні етапи процесу машинного навчання та його застосувань у різних галузях.

#### Тема 2. Попередня обробка даних

Основи методів попередньої обробки даних, включно з нормалізацією, стандартизацією та очищенням даних. Аналіз важливості якісної попередньої обробки для покращення результатів моделювання.

#### Тема 3. Регресійний аналіз

Лінійна та поліноміальна регресії. Регуляризація як метод боротьби з перенавчанням моделі.

#### Тема 4. Класифікація

Основи класифікації з використанням логістичної регресії та дерев рішень. Аналіз методів оцінки ефективності класифікаційних моделей.

#### Тема 5. Методи зниження розмірності

Огляд аналізу головних компонент (PCA) та лінійного дискримінантного аналізу (LDA) як методів зниження розмірності. Вплив зниження розмірності на продуктивність моделі.

#### Тема 6. Кластеризація

Метод k-середніх та ієрархічної кластеризації. Методи визначення оптимальної кількості кластерів.

#### Тема 7. Ансамблеві методи

Огляд основних ансамблевих методів, таких як беггінг, бустінг та випадковий ліс (Random Forest). Аналіз переваг використання ансамблів для покращення точності моделей.

#### Тема 8. Метод опорних векторів (SVM)

Принцип роботи SVM та використання ядрових функцій для роботи з нелінійними даними. Особливості застосування SVM у класифікації та регресії.

#### Тема 9. Навчання без вчителя

Аналіз основних методів навчання без вчителя, включаючи кластеризацію та асоціативні правила. Приклади застосування навчання без вчителя.

#### Тема 10. Оцінка моделі та налаштування гіперпараметрів

Методи оцінки моделі, включаючи перехресну перевірку та AUC-ROC. Розгляд стратегій налаштування гіперпараметрів для оптимізації продуктивності моделі.

#### Тема 11. Рекомендаційні системи

Основні типи рекомендаційних систем та їх застосування. Аналіз методів колаборативної фільтрації та фільтрації на основі змісту.

#### Тема 12. Аналіз часових рядів

Основи аналізу часових рядів та прогнозування. Моделі ARIMA та сезонної корекції.

#### Тема 13. Текстові дані та обробка природної мови

Методи обробки текстових даних, включаючи векторизацію тексту та аналіз настроїв. Застосування NLP у машинному навчанні.

#### Тема 14. Етика у машинному навчанні

Аналіз етичних проблем, пов'язаних з використанням машинного навчання. Розгляд питань приватності, упередженості та контролю за AI.

#### Тема 15. Вибіркове навчання

Огляд методів вибіркового навчання та їх застосування в умовах обмеженого навчального датасету. Розгляд стратегій активного навчання для ефективного використання даних.

#### Тема 16. Оптимізація моделей машинного навчання

Вивчення методів оптимізації моделей, включаючи градієнтний спуск та його варіації. Аналіз способів підвищення ефективності та швидкості навчання моделей.

### Теми практичних занять

Не передбачено навчальним планом

### Теми лабораторних робіт

#### Лабораторна робота №1. Попередня обробка даних

Практичне застосування методів попередньої обробки даних, включаючи нормалізацію, стандартизацію та очищення даних. Виконання аналізу вихідних даних та їх підготовка для подальшого моделювання.

#### Лабораторна робота №2. Регресійний аналіз

Розробка та оцінка моделі лінійної регресії. Вивчення впливу регуляризації на якість моделі. Аналіз резидуалів та перевірка припущень лінійної регресії.

#### Лабораторна робота №3. Класифікація з використанням логістичної регресії

Реалізація моделі логістичної регресії для задачі бінарної класифікації. Оцінка ефективності моделі за допомогою матриці помилок, точності, повноти та інших метрик.

#### Лабораторна робота №4. Методи зниження розмірності

Застосування аналізу головних компонент (PCA) для зниження розмірності датасету. Аналіз впливу зниження розмірності на продуктивність класифікаційних або регресійних моделей.

#### Лабораторна робота №5. Кластеризація з використанням методу k-середніх

Реалізація алгоритму k-середніх для кластеризації даних. Визначення оптимальної кількості кластерів за допомогою методу ліктя або інших методів оцінки.

#### Лабораторна робота №6. Використання ансамблевих методів для класифікації

Побудова моделей випадкового лісу (Random Forest) та градієнтного бустінгу для задачі класифікації. Аналіз впливу кількості дерев на точність моделі.

#### **Лабораторна робота №7. Моделювання з використанням методу опорних векторів (SVM)**

Реалізація класифікатора на основі SVM. Експериментування з різними ядровими функціями та параметрами регуляризації для оптимізації моделі.

#### **Лабораторна робота №8. Оцінка моделі та налаштування гіперпараметрів**

Використання технік перехресної перевірки для оцінки моделей. Практичне застосування методів налаштування гіперпараметрів, таких як Grid Search та Random Search, для покращення продуктивності моделей машинного навчання.

### **Самостійна робота**

*Індивідуальне завдання* передбачає розробку програмного рішення, яке демонструє застосування методів обробки природної мови (NLP) для аналізу та обробки текстових даних. Студент повинен проаналізувати задачу, визначити необхідні для її вирішення алгоритми обробки тексту та розробити програму, яка ефективно виконує поставлені задачі. Програма має включати в себе етапи попередньої обробки тексту, векторизації, а також використання моделей машинного навчання для аналізу тексту. Особливу увагу слід приділити якості обробки текстових даних, а також точності та надійності отриманих результатів. Важливим є розробка чіткого та добре структурованого коду.

Можливі теми для індивідуальних проєктів:

1. Система аналізу настроїв відгуків користувачів
2. Інструмент для автоматичного резюмування текстів
3. Класифікація текстових документів за темами
4. Система виявлення та виправлення помилок у текстах.

### **Література та навчальні матеріали**

1. Т.М. Басюк, В.В. Литвин, Л.М. Захарія, Н.Е. Кунанець. Машинне навчання: Навчальний посібник призначений для студентів, що навчаються за першим (бакалаврським) рівнем вищої освіти за спеціальностями галузі знань 12 „Інформаційні технології”. – Львів: Видавництво «Новий Світ - 2000», 2019. - 335 с.
2. Шайлендра Кадре. Machine Learning and Deep Learning Using Python and TensorFlow. – 2021. - 560 p.
3. Aurelien Geron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 3rd Edition. - 2022. – 850 p.
4. Jonathon Manning, Tim Nugent, Paris Buttfield-Addison, Mars Buttfield-Addison. Practical Simulations for Machine Learning. Using Synthetic Data for AI. – 2022. – 332 p.

## Система оцінювання

### Критерії оцінювання успішності студента та розподіл балів

Семестровий контроль проводиться в усній формі за екзаменаційними білетами. Результати поточного контролю враховуються як допоміжна інформація для виставлення оцінки з даної дисципліни. Студент вважається допущеним до семестрового екзамену за умови захисту усіх лабораторних робіт та індивідуального завдання. Бали нараховуються наступним чином:  
- екзамен – 100% семестрової оцінки

### Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

## Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

## Погодження

Силабус погоджено

28.08.2023

Завідувач кафедри  
Дмитро БРЕСЛАВСЬКИЙ

28.08.2023

Гарант ОП  
Оксана ТАТАРІНОВА