



## ОСНОВИ НАУКИ ПРО ДАНІ

### Робоча програма навчальної дисципліни (Силабус)

#### Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Перший (бакалаврський)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>123 Комп'ютерна інженерія</i>
Освітня програма	<i>Комп'ютерних системи та мережі</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>Очна (денна)</i>
Рік підготовки, семестр	<i>4 курс, весняний семестр</i>
Обсяг дисципліни	<i>4 кредитів, 120 годин</i>
Семестровий контроль/ контрольні заходи	<i>Залік</i>
Розклад занять	<i>// rozklad.kpi.ua</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: д.т.н, професор, Новотарський Михайло Анатолійович novotar@gmail.com, http://novotarsky.pp.ua Лабораторні: д.т.н, професор, Новотарський Михайло Анатолійович novotar@gmail.com, http://novotarsky.pp.ua</i>
Розміщення курсу	<i>https://ecampus.kpi.ua</i>

#### Програма навчальної дисципліни

##### 1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Навчальна дисципліна «Основи науки про дані» призначена для надання базових знань у галузі аналізу даних для розв'язування наукових та інженерно-технічних задач та програмування з врахуванням особливостей сучасних підходів та методів добування даних.

Навчальна дисципліна «Основи науки про дані» є вибірковою дисципліною.

В рамках даного курсу розглядається основне поняття про дані та їх властивості, обговорюються основні різновидності даних та їх атрибутів. Коротко наводиться інформація про способи представлення даних.

Далі розглядається проблема аналізу даних для отримання корисних знань. Розглянуто основні етапи аналізу даних та методи, які в даному випадку використовуються. Детально розглянуто такі задачі добування даних, як класифікацію та кластеризацію, а також задачі прогнозування та візуалізації.

Курс «Основи науки про дані» включає ознайомлення студентів з основними етапами добування даних та їх реалізацією. На початковому етапі розглянуто принципи початкової підготовки даних та побудови моделі, а на наступних етапах – принципи очистки даних та аналізу отриманих результатів.

Метою вивчення курсу «Основи науки про дані» є вивчення сучасних методів та технологій добування даних, підготовка студентів у виборі та використанні методів початкової підготовки, очистки та побудови моделей.

**Предметом** дисципліни є:

- методи аналізу даних;
- методи очистки даних;
- методи побудови моделей.

Згідно з вимогами освітньої програми студенти після засвоєння навчальної дисципліни мають продемонструвати такі результати навчання:

**Знання:** основних методів початкової підготовки даних та відповідних ефективних алгоритмів розв'язування цих задач на ПК; знання методів класифікації та кластеризації для обробки даних різного обсягу, знання методів прогнозування та візуалізації; знання процесу добування даних, який використовується у сучасних системах добування даних.

**Вміння** вибирати та обґрунтовувати використання на практиці тих чи інших методів добування даних, стійких до похибок і найбільш ефективних при їх практичній реалізації на ПК, вміння проводити верифікацію отриманих результатів.

**Досвід:** студент повинен знати основні принципи розробки алгоритмів та програмного забезпечення розв'язування задач добування даних на ПК; досліджувати алгоритми добування даних, виявляти їх переваги та недоліки, обирати оптимальні алгоритми розв'язування задач, обробки даних та розробляти програми розв'язування задач; виконувати аналіз і опрацювання результатів розв'язування задач, використовувати методи оптимізації.

## **2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)**

Матеріал навчальної дисципліни пов'язаний з матеріалами, які вивчалися у курсах «Вища математика», «Програмування» та «Дискретна математика».

Знання та практичні навички, які отримані в рамках даної навчальної дисципліни, можуть бути застосовані при вивченні наступних курсів: «Комп'ютерне моделювання», «Основи програмної інженерії», «Системне програмування», «Технологія розподілених обчислень», «Мережні технології» та інші.

## **3. Зміст навчальної дисципліни**

Навчальна дисципліна «Основи науки про дані» включає вивчення наступних тем.

**Тема 1. Вступ. Базові відомості про дані.** Поняття даних. Набір даних та їх атрибутів. Вимірювання. Шкали. Типи наборів даних: записи, хімічні дані, графи. Формати зберігання даних. Основні положення про бази даних. Системи управління базами даних. Класифікація видів даних. Метадані.

**Тема 2. Основні положення аналізу даних.** Поняття про процес добування даних, процес **KDD** (Knowledge Discovery in Databases). Математична база методів добування даних, статистичні методи. Причини виникнення та мотивація розвитку методів добування даних. Суть технології добування даних, мультидисциплінарність. Порівняння методів машинного навчання та методів

добування даних. Проблеми підготовки даних. Огляд алгоритмів та методів попередньої обробки даних: очистка, перетворення та скорочення.

**Тема 3. Базові поняття про бібліотеку Pandas. Структура даних Pandas Series.** Основні характеристики бібліотеки. Особливості інсталяції під різними операційними системами. Переваги структури Series у порівнянні з масивами NumPy. Визначення масиву міток за замовчуванням та у вигляді рядків. Способи виводу даних типу Series. Доступ до елементів Series та їх модифікація. Створення Series з використанням масиву NumPy. Динамічне копіювання посилань на елементи при створенні Series. Створення Series з словника. Фільтрація елементів Series. Групові математичні операції Series з скалярними величинами. Операції  $/$ ,  $*$ ,  $-$ ,  $+$  з двома об'єктами Series. Робота з повторюваними елементами в Series. Використання NaN: перевірка наявності значень NaN, фільтрація елементів з NaN

**Тема 4. Основні характеристики та способи застосування структури даних DataFrame.** Переваги структури DataFrame у порівнянні з двовимірними масивами NumPy. Створення DataFrame за замовчуванням та з використанням словника. Створення міток рядків у DataFrame. Генерування DataFrame з використанням методу `reshape()`. Перетворення DataFrame у NumPy масив. Зчитування стовпця DataFrame у Series. Зчитування рядків DataFrame. Додавання стовпця з однаковими значеннями. Оновлення значень у стовпці. Модифікація значення елемента та перевірка його входження. Фільтрація у структурі DataFrame. Повторення даних. Арифметичні методи `add()`, `sub()`, `div()`, `mul()`. Арифметичні операції між DataFrame і Series. Проблема відображення даних DataFrame: поворот структури даних та приклади його застосування. Поворот з одним та кількома стовпцями. Типові помилки при використанні методу повороту. Методи `stack()` і `unstack()` та приклади їх застосування. Застосування функції `melt()`, пропуск стовпців при використанні `melt()`.

**Тема 5. Функції бібліотеки Pandas. Модифікація великих наборів даних.** Функції округлення: `trunc()`, `around()`, `floor()`, `ceil()`. Функції взяття логарифмів: `log2()`, `log10()`, `log()`. Функції обчислення суми: `sum()`, `cumsum()`. Функції обчислення добутку: `prod()`, `cumprod()`. Функція обчислення скінченних різниць: `diff()`. Знаходження найменшого спільного кратного (LCM - Lowest Common Multiple). Знаходження найбільшого спільника дільника (GCD - Greatest Common Denominator). Тригонометричні функції та функції множин. Функції для стовпців. Функція статистики. Сортування Series. Сортування DataFrame. Кореляція і коваріантність у Series та DataFrame. Інструменти I/O у Pandas: робота з файлами різних типів. Перегляд фрагментів набору даних. Визначення типів даних колонок. Роздруківка технічної інформації про набір даних. Визначення кількості NaN елементів у наборі даних. Заміна відсутніх даних константним значенням. Заповнення середнім значенням по стовпцю. Заповнення інтерпольованими даними. Видалення рядків та стовпців з пропусками. Встановлення порогу пропусків. Розбиття на відрізки однакового розміру. Злиття двох структур DataFrame.

**Тема 6. Індексція та реіндексція наборів даних.** Вирівнювання даних у Pandas. Зміна порядку виводу наявних даних з маркером відсутності. Метод `reindex()` для Series та для DataFrame. Реіндексція з ключовим словом `axis`. Використання індексції стороннього набору даних. Метод `align()`. Вирівнювання об'єктів з використанням `align(s,join="outer")`, `align(s,join="inner")`, `align(s,join="left")`, `align(s,join="right")`. Метод переіменування `rename(str.upper)`. Метод переіменування `rename()` з відображенням. Динамічне перейменування по осях.

**Тема 7. Багаторівнева індексція. Створення та використання об'єкта Multiindex.**

Створення об'єкта `MultiIndex` з використанням методів `from_tuples()`, `from_product()`, `from_frame()`. Автоматичний `MultiIndex`. Керування відображенням індексів набору даних. Використання кортежів, як міток. Вирівнювання даних з `MultiIndex`. Застосування `reindex()` для Series та DataFrame. Розширене індексування з ієрархічним індексом з методом `loc()`. Використання зрізів. Перестановка рівнів за допомогою `swarlevel`. Зміна порядку рівнів за допомогою `reorder_levels`. Метод `rename()` для перейменування `MultiIndex`. Перейменування основного індексу. Застосування методу `MultiIndex.set_names()`. Сортування `MultiIndex`. Сортування `MultiIndex` по рівнях.

Задавання level по номеру для параметру index та columns. Метод is\_monotonic\_increasing(). Зріз по сортованому MultiIndex. Метод take() для структур Series та DataFrame. Індексція за допомогою IntervalIndex. Дані типу Categoricals. Перетворення Series у тип category. Бінінг (створення контейнерів) даних за допомогою cut. Створення діапазонів інтервалів. Застосування діапазонів як IntervalIndex у DataFrame. Діапазони з параметром freq. Агрегування даних по єдиному стовпцю. Застосування методів агрегування у параметрі aggfunc. Метод множинної агрегації в pandas DataFrame. Використання різних агрегацій по стовпцях. Сортування зведеної таблиці

**Тема 8. Відображення великих наборів даних. Бібліотека Matplotlib.** Способи інсталяції Matplotlib. Відображення найпростішого графіка. Графік з кумулятивною сумою. Відображення DataFrame з використанням міток. Функціональна залежність одного стовпця від іншого. Графік типу Bar для структур Series та DataFrame. Графік типу «складений Bar» вертикальний. Графік типу «складений Bar» горизонтальний. Гістограма для структури DataFrame. Складена гістограма. Горизонтальна кумулятивна гістограма. Групові гістограми по стовпцях. Діаграми площ у стопку. Діаграми площ, які накладаються. Діаграма розсіювання (точкова діаграма). Багатоколірна мультидіаграма розсіювання по стовпцях. Кругова діаграма для DataFrame. Модифікація кругових діаграм. Матрична діаграма розсіювання. Графік щільності. Криві Ендрюса. Графік відставання. Графік автокореляції. Бутстреп графік. RadViz. Загальні аргументи стилю графіка. Керування легендою. Керування мітками. Логарифмічні шкали. Побудова графіка на вторинній осі ординат. Графік таблиця. Роздруківка списку кольорових стилів. Процес формування графіка з використанням стилю. Темний фон на графіку. Створення кількох графіків методом subplots(). Використання спільних осей Y та X.

**Тема 9. Задача логістичної регресії з використанням бібліотеки PyTorch.** Інсталяція PyTorch під Anzconda. Перевірка працездатності бібліотеки PyTorch. Основи положення роботи з PyTorch. Створення тензора з нульовим або одиничним значенням. Створення довільного тензора. Перетворення pandas Series у PyTorch tensor. Перетворення pandas DataFrame у PyTorch tensor. Стандартні набори даних. Набор даних MNIST. Завантаження MNIST. Перетворення формату рисунка. Підготовка даних до навчання за допомогою DataLoaders. Теоретичні основи логістичної регресії. Прямий хід логістичної регресії. Початкова ініціалізація параметрів моделі. Обчислення у з урахуванням градієнтного спуску для параметрів. Застосування softmax(). Два варіанти обчислення softmax(). Крос-ентропійна функція втрат. Два варіанти обчислення крос-ентропійної втрати. Зворотний хід логістичної регресії. Тренування моделі. Перевірка точності. API вищого рівня логістичної моделі. Об'єктно-орієнтована рефакторизація. Використання torch.nn. Використання torch.nn.Module.

**Тема 10. Створення глибокої нейронної мережі з використанням PyTorch.** Навчальний набір Fashion MNIST. Код завантаження FashionMNIST. Ітерування і візуалізація набору даних Fashion-MNIST. Створення власного набору даних Custom Dataset. Методи \_\_init\_\_, \_\_len\_\_, .\_\_getitem\_\_. Код підготовки даних у DataLoader. Ітерування у DataLoader. Перетворення даних. Побудова нейронної мережі. Визначення класу NeuralNetwork. Використання моделі. Шари моделі. Диференціювання з використанням torch.autograd. Тензори, функції та обчислювальний графік. Обчислювальні градієнти. Вимкнення відстеження градієнта. Оптимізація модельних параметрів. Гіперпараметри. Цикл оптимізації. Функція втрат. Способи оптимізації моделі.

Навчальні матеріали та ресурси

**Базова:**

1. Новотарський М.А. Лекції х курсу «Основи науки про дані» // <https://cloud.comsys.kpi.ua/s/pjw4nAstZFPsDos>
2. Новотарський М.А. Методичні вказівки до виконання лабораторних робіт з курсу «Основи науки про дані» // Новотарський М.А. Лекції х курсу «Основи науки про дані» //
3. Чубакова И. Data Mining // Курс «ИНТУИТ» / <https://intuit.ru/studies/courses/6/6/info>

**Додаткова:**

1. Wang L., Fu X. Data Mining with Computational Intelligence. –Springer, 2005. –280 p.

**Навчальний контент****4. Методика опанування навчальної дисципліни (освітнього компонента)**

Структура навчальної дисципліни «Основи науки про дані» в таблиці 1 .

Таблиця 1

Структура навчальної дисципліни «Основи науки про дані»

Назва розділів і тем	Кількість годин			
	Всього	у тому числі		
		Лекції	Лабораторні (комп'ютерний практикум)	СРС
<b>Тема 1. Вступ. Базові відомості про дані.</b>	8	2		6
<b>Тема 2. Основні положення аналізу даних.</b>	8	2	2	4
<b>Тема 3. Базові поняття про бібліотеку Pandas. Структура даних Pandas Series.</b>	12	4	2	6
<b>Тема 4. Основні характеристики та способи застосування структури даних DataFrame.</b>	20	4	2	14
<b>Тема 5. Функції бібліотеки Pandas. Модифікація великих наборів даних.</b>	17	4	2	11
<b>Тема 6. Індексція та реіндексція наборів даних.</b>	16	2	2	14
<b>Тема 7. Багаторівнева індексція. Створення та використання об'єкта Multiindex.</b>	18	2	2	14
<b>Тема 8. Відображення великих наборів даних. Бібліотека Matplotlib.</b>	26	2	2	22
<b>Тема 9. Задача логістичної регресії з використанням бібліотеки PyTorch.</b>	24	5	2	17
<b>Тема 10. Створення глибокої нейронної мережі з використанням PyTorch.</b>	24	5	2	17
<b>Семестрова контрольна робота</b>	8	2	0	6
<b>Залік</b>	<b>14</b>	<b>2</b>	<b>0</b>	<b>12</b>
<b>Всього в семестрі:</b>	<b>120</b>	<b>36</b>	<b>18</b>	<b>66</b>

Тематика лекційних занять, яка сформована відповідно до тем, що розглядаються у рамках навчальної дисципліни, наведена у таблиці 2.

Таблиця 2

Лекційні заняття

№ з/п	Назва теми лекції та перелік основних питань
1	<p><b>Вступ. Базові відомості про дані.</b> Поняття даних. Набір даних та їх атрибутів. Вимірювання. Шкали. Типи наборів даних: записи, хімічні дані, графи. Формати зберігання даних. Основні положення про бази даних. Системи управління базами даних. Класифікація видів даних. Метадані.</p>
2	<p><b>Основні положення аналізу даних.</b> Поняття про процес добування даних, процес <b>KDD</b> (Knowledge Discovery in Databases). Математична база методів добування даних, статистичні методи. Причини виникнення та мотивація розвитку методів добування даних. Суть технології добування даних, мультидисциплінарність. Порівняння методів машинного навчання та методів добування даних. Проблеми підготовки даних. Огляд алгоритмів та методів попередньої обробки даних: очистка, перетворення та скорочення. Закономірності в даних: неочевидні закономірності, тобто закономірності, які є не можуть бути знайдені стандартними методами обробки інформації або експертним шляхом. Об'єктивні закономірності, тобто закономірності, які повністю відповідають дійсності, на відміну від експертної думки, яке завжди є суб'єктивною. Практично корисні закономірності, тобто закономірності, що мають конкретне значення, якому можна знайти практичне застосування.</p>
3	<p><b>Базові поняття про бібліотеку Pandas.</b> Основні характеристики бібліотеки. Особливості інсталяції під різними операційними системами. Особливості структур даних Pandas. Одновимірні та двовимірні структури даних. Огляд функціональних можливостей бібліотеки щодо попередньої обробки даних. Способи попередньої обробки даних, пропущені та дубльовані дані.</p>
4	<p><b>Структура даних Pandas Series.</b> Основні характеристики бібліотеки. Особливості інсталяції під різними операційними системами. Переваги структури Series у порівнянні з масивами NumPy. Визначення масиву міток за замовчуванням та у вигляді рядків. Способи виводу даних типу Series. Доступ до елементів Series та їх модифікація. Створення Series з використанням масиву NumPy. Динамічне копіювання посилань на елементи при створенні Series. Створення Series з словника. Фільтрація елементів Series. Групові математичні операції Series з скалярними величинами. Операції /, *, -, + з двома об'єктами Series. Робота з повторюваними елементами в Series. Використання NaN: перевірка наявності значень NaN, фільтрація елементів з NaN.</p>
5	<p><b>Створення та модифікація структури даних DataFrame.</b> Переваги структури DataFrame у порівнянні з двовимірними масивами NumPy. Створення DataFrame за замовчуванням та з використанням словника. Створення міток рядків у DataFrame. Генерування DataFrame з використанням методу reshape(). Перетворення DataFrame у NumPy масив. Зчитування стовпця DataFrame у Series. Зчитування рядків DataFrame. Додавання стовпця з однаковими значенням. Оновлення значень у стовпці. Модифікація значення елемента та перевірка його входження.</p>
6	<p><b>Основні методи структури даних DataFrame.</b> Фільтрація у структурі DataFrame. Повторення даних. Арифметичні методи add(), sub(), div(), mul(). Арифметичні операції між DataFrame і</p>

№ з/п	Назва теми лекції та перелік основних питань
	Series. Проблема відображення даних DataFrame: поворот структури даних та приклади його застосування. Поворот з одним та кількома стовпцями. Типові помилки при використанні методу повороту. Методи stack() і unstack() та приклади їх застосування. Застосування функції melt(), пропуск стовпців при використанні melt().
7	<b>Функції бібліотеки Pandas.</b> Функції округлення: trunc(), around(), floor(), ceil(). Функції взяття логарифмів: log2(), log10(), log(). Функції обчислення суми: sum(), cumsum(). Функції обчислення добутку: prod(), cumprod(). Функція обчислення скінчених різниць: diff(). Знаходження найменшого спільного кратного (LCM - Lowest Common Multiple). Знаходження найбільшого спільника дільника (GCD - Greatest Common Denominator). Тригонометричні функції та функції множин. Функції для стовпців. Функція статистики.
8	<b>Модифікація великих наборів даних.</b> Сортування Series. Сортування DataFrame. Кореляція і коваріантність у Series та DataFrame. Інструменти I/O у Pandas: робота з файлами різних типів. Перегляд фрагментів набору даних. Визначення типів даних колонок. Роздрукування технічної інформації про набір даних. Визначення кількості NaN елементів у наборі даних. Заміна відсутніх даних константним значенням. Заповнення середнім значенням по стовпцю. Заповнення інтерпольованими даними. Видалення рядків та стовпців з пропусками. Встановлення порогу пропусків. Розбиття на відрізки однакового розміру. Злиття двох структур DataFrame.
9	<b>Реіндексація наборів даних.</b> Вирівнювання даних у Pandas. Зміна порядку виводу наявних даних з маркером відсутності. Метод reindex() для Series та для DataFrame. Реіндексація з ключовим словом axis. Використання індексації стороннього набору даних. Метод align(). Вирівнювання об'єктів з використанням align(s,join="outer"), align(s,join="inner"), align(s,join="left"), align(s,join="right"). Метод переіменування rename(str.upper). Метод переіменування rename() з відображенням. Динамічне перейменування по осях.
10	<b>Багаторівнева індексація. Створення MultiIndex.</b> Створення об'єкта MultiIndex з використанням методів from_tuples(), from_product(), from_frame(). Автоматичний MultiIndex. Керування відображенням індексів набору даних. Використання кортежів, як міток. Вирівнювання даних з MultiIndex. Застосування reindex() для Series та DataFrame. Розширене індексування з ієрархічним індексом з методом loc(). Використання зрізів.
11	<b>Методи об'єкта Multiindex.</b> Перестановка рівнів за допомогою swaplevel. Зміна порядку рівнів за допомогою reorder_levels. Метод rename() для перейменування MultiIndex. Перейменування основного індексу. Застосування методу MultiIndex.set_names(). Сортування MultiIndex. Сортування MultiIndex по рівнях. Задавання level по номеру для параметру index та columns. Метод is_monotonic_increasing(). Зріз по сортованому MultiIndex. Метод take() для структур Series та DataFrame. Індикація за допомогою IntervalIndex. Дані типу Categoricals. Перетворення Series у тип category. Бінінг (створення контейнерів) даних за допомогою cut. Створення діапазонів інтервалів. Застосування діапазонів як IntervalIndex у DataFrame. Діапазони з параметром freq. Агрегування даних по єдиному стовпцю. Застосування методів агрегування у параметрі aggfunc. Метод множинної агрегації в pandas DataFrame. Використання різних агрегацій по стовпцях. Сортування зведеної таблиці
12	<b>Відображення великих наборів даних. Бібліотека Matplotlib.</b> Способи інсталяції Matplotlib. Відображення найпростішого графіка. Графік з кумулятивною сумою. Відображення DataFrame з використанням міток. Функціональна залежність одного стовпця від іншого. Графік типу Bar для структур Series та DataFrame. Графік типу «складений Bar» вертикальний. Графік типу «складений Bar» горизонтальний. Гістограма для структури



№ з/п	Назва теми лекції та перелік основних питань
	DataFrame. Складена гістограма. Горизонтальна кумулятивна гістограма. Групові гістограми по стовпцях. Діаграми площ у стопку. Діаграми площ, які накладаються. Діаграма розсіювання (точкова діаграма).
13	<b>Діаграми у Matplotlib.</b> Багатоколірна мультидіаграма розсіювання по стовпцях. Кругова діаграма для DataFrame. Модифікація кругових діаграм. Матрична діаграма розсіювання. Графік щільності. Криві Ендрюса. Графік відставання. Графік автокореляції. Бутстреп графік. RadViz. Загальні аргументи стилю графіка. Керування легендою. Керування мітками. Логарифмічні шкали. Побудова графіка на вторинній осі ординат. Графік таблиця. Роздруківка списку кольорових стилів. Процес формування графіка з використанням стилю. Темний фон на графіку. Створення кількох графіків методом subplots(). Використання спільних осей Y та X.
14	<b>Задача логістичної регресії з використанням бібліотеки PyTorch.</b> Інсталяція PyTorch під Anzconda. Перевірка працездатності бібліотеки PyTorch. Основи положення роботи з PyTorch. Створення тензора з нульовим або одиничним значенням. Створення довільного тензора. Перетворення pandas Series у PyTorch tensor. Перетворення pandas DataFrame у PyTorch tensor. Стандартні набори даних.
15	<b>Набір даних MNIST та робота з ним.</b> Завантаження MNIST. Перетворення формату рисунка. Підготовка даних до навчання за допомогою DataLoaders. Теоретичні основи логістичної регресії. Прямий хід логістичної регресії. Початкова ініціалізація параметрів моделі. Обчислення $u$ з урахуванням градієнтного спуску для параметрів. Застосування softmax(). Два варіанти обчислення softmax(). Крос-ентропійна функція втрат. Два варіанти обчислення крос-ентропійної втрати. Зворотний хід логістичної регресії. Тренування моделі. Перевірка точності. API вищого рівня логістичної моделі. Об'єктно-орієнтована рефакторизація. Використання torch.nn. Використання torch.nn.Module.
16	<b>Створення глибокої нейронної мережі з використанням PyTorch.</b> Навчальний набір Fashion MNIST. Код завантаження FashionMNIST. Ітерування і візуалізація набору даних Fashion-MNIST. Створення власного набору даних Custom Dataset. Методи <code>__init__</code> , <code>__getitem__</code> . Код підготовки даних у DataLoader. Ітерування у DataLoader. Перетворення даних. Побудова нейронної мережі. Визначення класу NeuralNetwork. Використання моделі. Шари моделі. Диференціювання з використанням torch.autograd. Тензори, функції та обчислювальний графік. Обчислювальні градієнти. Вимкнення відстеження градієнта. Оптимізація модельних параметрів. Гіперпараметри. Цикл оптимізації. Функція втрат. Способи оптимізації моделі.
17	<b>Семестрова контрольна робота</b>
18	<b>Залік</b>

## 5. Самостійна робота студента

Самостійна робота студента складається з теоретичної та практичної складової. Теоретична складова передбачає вивчення додаткового матеріалу, що поглиблює знання, які отримані на лекції. Матеріал, який необхідно додатково вивчити, та літературні джерела для вивчення даного матеріалу вказані у таблиці 4. Практична складова самостійної роботи студента полягає у виконанні лабораторних робіт, перелік яких наведено в таблиці 3.

Основне завдання циклу лабораторних занять полягає у набутті студентами необхідних практичних навичок розробки алгоритмів та програмного забезпечення для розв'язування задач обробки даних на ПК.



## Перелік лабораторних робіт

№ з/п	Назва лабораторної роботи (комп'ютерного практикуму)	Кількість ауд. годин
1	Лабораторна робота № 1. Підсумкові функції та відображення Тема 2, Тема 3.	4
2	Лабораторна робота № 2. Індексування, вибір, редагування набору даних Тема 4, Тема 5.	4
3	Лабораторна робота № 3. Робота з пропущеними даними Тема 6, Тема 7, Тема 8.	4
4	Лабораторна робота № 4. Розв'язування задачі лінійної регресії Тема 9, Тема 10	6

## Тематика самостійної роботи студента

№ з/п	Назва теми лекції та перелік основних питань (перелік дидактичних засобів, посилання на літературу та завдання на СРС)
1	<b>Вступ. Базові відомості про дані.</b> Види даних та способи їх обробки. Операції над даними. Класифікація інформації за видами і типами. Види кодувань. Структура даних та одиниці вимірювання. <i>Література</i> <a href="https://hi-news.pp.ua/kompyuteri/1379-vidi-danih-sposobi-yih-obrobki.html">https://hi-news.pp.ua/kompyuteri/1379-vidi-danih-sposobi-yih-obrobki.html</a>
2	<b>Основні положення аналізу даних.</b> Поняття про взаємозв'язок статистичних показників. Види та форми зв'язків між явищами. Прийоми виявлення щільності зв'язку між показниками досліджуваних явищ. <i>Література</i> <a href="https://xreferat.com/22/2260-1-osnovn-polozhennya-kompleksnogo-statistichnogo-anal-zu-danih-u-pravov-iy-statistic.html">https://xreferat.com/22/2260-1-osnovn-polozhennya-kompleksnogo-statistichnogo-anal-zu-danih-u-pravov-iy-statistic.html</a>
3	<b>Базові поняття про бібліотеку Pandas.</b> Загальні поняття про Pandas. Сфера застосувань бібліотеки Pandas. Додаткові властивості структур даних у Pandas/ <i>Література</i> <a href="https://coderlessons.com/tutorials/mashinnoe-obuchenie/uchebnik-tensorflow/6-uchebnik-po-python-pandas">https://coderlessons.com/tutorials/mashinnoe-obuchenie/uchebnik-tensorflow/6-uchebnik-po-python-pandas</a>
4	<b>Основні характеристики та способи застосування структури даних DataFrame.</b> class pandas.DataFrame. Атрибути та методи структури даних DataFrame <i>Література</i> <a href="https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html">https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html</a>
5	<b>Функції бібліотеки Pandas. Модифікація великих наборів даних.</b> Бінарні функції. Функції порівняння значень. Перевірка на еквівалентність. Порівняння масивів. Об'єднання даних,

№ з/п	Назва теми лекції та перелік основних питань (перелік дидактичних засобів, посилання на літературу та завдання на СРС)
	які перекриваються. Методи описової статистики. Підрахунок значень та мода значень. Дискретизація неперервних значень. <i>Література</i> <a href="https://pyprog.pro/pd/pd_11_basic_functionality.html">https://pyprog.pro/pd/pd_11_basic_functionality.html</a>
6	<b>Індексація та реіндексація наборів даних.</b> Різні варіанти індексації. Методи для індексації. <i>Література</i> <a href="https://coderlessons.com/tutorials/python-technologies/vyuchit-python-panda/python-pandas-indeksirovanie-i-vybor-dannykh">https://coderlessons.com/tutorials/python-technologies/vyuchit-python-panda/python-pandas-indeksirovanie-i-vybor-dannykh</a>
7	<b>Багаторівнева індексація. Створення та використання об'єкта Multiindex.</b> Ієрархічне індексування. Створення об'єкта MultiIndex (ієрархічний індекс). Відновлення міток рівнів. Базове індексування на осі за допомогою MultiIndex. Визначені рівні. Розширене індексування з ієрархічним індексом. Використання слайсерів. Поперечний переріз. <i>Література</i> <a href="https://pandas.pydata.org/docs/user_guide/advanced.html">https://pandas.pydata.org/docs/user_guide/advanced.html</a>
8	<b>Відображення великих наборів даних. Бібліотека Matplotlib.</b> Оформлення графіків. Розташування графіків. Робота з осями. Інтерактивні можливості. <i>Література</i> <a href="https://jenyay.net/Matplotlib/Matplotlib">https://jenyay.net/Matplotlib/Matplotlib</a>
9	<b>Задача логістичної регресії з використанням бібліотеки PyTorch.</b> Установка PyTorch. Підготовка даних в PyTorch. Навчання моделі за допомогою PyTorch. Розгортання моделі за допомогою Windows. <i>Література</i> <a href="http://cloud-5.bitp.kiev.ua/?page_id=605">http://cloud-5.bitp.kiev.ua/?page_id=605</a> <a href="https://docs.microsoft.com/ru-ru/windows/ai/windows-ml/tutorials/pytorch-installation">https://docs.microsoft.com/ru-ru/windows/ai/windows-ml/tutorials/pytorch-installation</a>
10	<b>Створення глибокої нейронної мережі з використанням PyTorch.</b> Класифікація зображень за допомогою PyTorch. Проблема класифікації. Завантаження даних PyTorch. Створення навчального набору даних. Волідація та контрольні набори даних. <i>Література</i> <a href="https://habr.com/ru/company/piter/blog/512158/">https://habr.com/ru/company/piter/blog/512158/</a>

## Політика та контроль

### 6. Політика навчальної дисципліни (освітнього компонента)

Під час занять з навчальної дисципліни «Основи науки про дані» студенти повинні дотримуватись певних дисциплінарних правил:

- забороняється запізнюватись на заняття;
- при вході викладача, на знак привітання, особи, які навчаються в КПІ ім. Ігоря Сікорського, повинні встати;
- не допускаються сторонні розмови або інший шум, що заважає проведенню занять;
- виходити з аудиторії під час заняття допускається лише з дозволу викладача;
- не допускається користування мобільними телефонами та іншими технічними засобами без дозволу викладача.

Лабораторні роботи здаються особисто з попередньою перевіркою теоретичних знань, які необхідні для виконання лабораторної роботи. Перевірка практичних результатів включає перевірку коду та виконання тестових завдань.

В процесі навчання викладач має право нарахувати до 5 заохочувальних балів за дострокове виконання лабораторної роботи, за проявлений творчий підхід при виконанні індивідуального завдання або за активну участь у обговоренні питань, що пов'язані з тематикою лекції або практичного заняття.

За виконання та здачу лабораторної роботи після зазначеного дедлайну, за значну кількість пропущених занять, або за порушення правил поведінки на заняттях викладач може призначити до 5 штрафних балів.

При проведенні контрольних заходів та при виконанні лабораторних робіт студенти мають дотримуватися правил академічної доброчесності. При виявленні значного відсотку списування або плагіату викладач може відмовити у прийнятті даної роботи та вимагати доброчесного виконання навчального плану.

## 7. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Види контролю з навчальної дисципліни включають:

### **Лабораторні роботи:**

Заплановано самостійне виконання 4 лабораторних робіт.

Теми лабораторних робіт узгоджені у часі та за змістом з темами лекцій. Виконання лабораторних робіт у повному обсязі дозволяє набути практичних навичок застосування чисельних методів до розв'язування різних математичних задач та оволодіти сучасними технологіями програмування алгоритмів, які побудовані на основі даних методів.

### **Поточний контроль:**

Передбачено 5 поточних тестувань закритими тестами у системі TCXAM, які повністю охоплюють тематику даної навчальної дисципліни. Кожний поточний закритий тест містить 10 питань та триває 10 хв. Загальний час на проведення становить 2 години та включає час тестування та час на вирішення організаційних питань. У випадку дистанційного навчання закритий поточний тест проводиться на початку лекції, яка слідує за лекцією, що завершує чергову тему. При очному навчанні час чергового поточного тестування призначається викладачем за узгодженням зі студентами.

### **Семестровий контроль:**

Семестровий закритий тест проводиться в кінці семестру, триває 30 хвилин та складається з 25 питань.

### **Залік:**

проводиться у вигляді співбесіди зі студентом для об'єктивного визначення рівня знань, умінь та практичних навичок, отриманих за семестр

Оскільки кредитний модуль має семестрову атестацію у вигляді заліку, рейтингова система оцінювання побудована за типом PCO – 1. Рейтинг студента з кредитного модуля складається з балів, які він отримує за види робіт відповідно до таблиці 5.

Таблиця 5

Оцінювання окремих видів навчальної роботи студента

4 курс	
Вид навчальної роботи	Максимальна кількість балів
Виконання та захист лабораторної роботи №1	10
Виконання та захист лабораторної роботи №2	10
Виконання та захист лабораторної роботи №3	10

Виконання та захист лабораторної роботи № 4	10
Поточний контроль	30
Семестровий контроль закритими тестами	30
<b>Усього за семестр</b>	<b>100</b>

Індивідуальний семестровий рейтинг (**RD**) студента з навчальної дисципліни складається з балів, які він отримує за:

- 1) Поточний контроль: тематичні тестування закритими тестами у системі ТСЕХАМ (по 10 хвилин) –  $6 \times 5 = 30$  балів.
- 2) Семестровий контроль, який включає тестування закритими тестами у системі ТСЕХАМ (60 хвилин) – 30 балів,

Разом за семестровий контроль та поточний контроль – 60 балів.

- 3) Виконання лабораторних робіт.

*Протягом семестру студенти виконують 4 лабораторних роботи.*

Максимальна кількість балів за кожну лабораторну роботу – 10.

Бали нараховуються за:

- своєчасність подання роботи до захисту 0 – 1 бал,
- оформлення протоколу лабораторної роботи 0 – 1 бал,
- виконання змістовного завдання на роботу 0 – 5 балів,
- перевірка теоретичних знань, необхідних для виконання лабораторної роботи 0 – 5 балів.

Разом за лабораторні роботи (максимальна кількість балів) – 40.

*Виконання семестрового контрольного тестування.*

Семестровий контрольний тест виконується на кафедральному сервері у системі тестування «ТСЕХАМ».

Випадковим чином з кожної теми вибирається однакова кількість питань, загальним числом 30 питань, 4 варіанти відповідей на кожне питання, час виконання тесту – 60 хвилин.

*Виконання поточних контрольних тестів.* Поточні контрольні тести виконуються на кафедральному сервері у системі тестування «ТСЕХАМ».

Кожний поточний контрольний тест складається з 10 запитань.

Повна ґрунтовна відповідь на запитання тесту 0.5 - бала. Відповідь, яка включає недолік неповноти – 0.375 бала.

Відповідь, що містить дві вибрані неправильні відповіді - 0.25 бала.

Відповідь, що містить дві вибрані неправильні відповіді - 0.125 бала.

Відповідь, що не містить правильних варіантів відповіді - 0 балів.

Розрахунок розміру шкали (R) рейтингу:

Сума вагових балів контрольних заходів протягом семестру становить:

$$R = \sum_k r_k, \text{ де } r_k \text{ максимальний рейтинговий бал кожного з контрольних заходів (контрольне}$$

тестування, лабораторні роботи). Розмір рейтингової шкали з кредитного модуля становить:

$$R = 60 + 40 = 100 \text{ балів.}$$

У разі пропуску занять студентом без поважної причини нараховуються штрафні санкції у вигляді 1 бала з загальної суми балів за 1 годину пропуску (але не більше, ніж 0,1 R).

Індивідуальний семестровий рейтинг студента (підсумкова семестрова рейтингова оцінка **RD**) є сумою балів, отриманих студентом протягом семестру за участі у передбачених контрольних заходах (контрольні та лабораторні роботи).

Необхідною умовою допуску студента до заліку є його індивідуальний семестровий рейтинг (**RD**) не менший, ніж 30 балів, відсутність повної заборгованості з лабораторних робіт та не менше, ніж одна позитивна атестація. За невиконання хоча б однієї зі згаданих вимог студент до заліку не допускається.

Сума підсумкової семестрової (RD) та залікової рейтингових оцінок у балах становить підсумкову семестрову рейтингову оцінку, яка перераховується в оцінки за національною шкалою та шкалою ECTS (табл. 6).

Таблиця 6

Відповідність рейтингових балів оцінкам за університетською шкалою

<i>Кількість балів</i>	<i>Оцінка</i>
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

#### **8. Додаткова інформація з дисципліни (освітнього компонента)**

Викладання дисципліни «Алгоритми та методи обчислень» для спеціальності «Комп'ютерна інженерія» має свою специфіку, яка пов'язана з тим, що розробка та експлуатація комп'ютерної техніки потребує знання правил побудови та аналізу алгоритмів. Знайомство з універсальними моделями алгоритмів дозволяє визначити складність задачі та можливість її розв'язання за допомогою комп'ютера. Значна увага повинна приділятися підходам, що сприяють засвоєнню теоретичного матеріалу та практичних методів алгоритмізації, обчислення математичних задач, вивченню особливостей застосування методів, створення високоефективних алгоритмів реалізації чисельних методів. Саме чисельні методи лежать в основі всіх розрахунків, які можна виконати з використанням комп'ютерів.

#### **Робочу програму навчальної дисципліни (силабус):**

Складено д.т.н., проф. Новотарський Михайло Анатолійович

Ухвалено кафедрою ОТ (протокол № 10 від 25.05.22)

Погоджено Методичною комісією ФІОТ (протокол № 10 від 9.06.2)

...