



Національний технічний
університет України «КИЇВСЬКИЙ
ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені
ІГОРЯ СІКОРСЬКОГО»



Кафедра
обчислювальної
техніки



Sigma Software
University

**Освітній компонент сертифікатної програми
«Технології аналізу та обробки даних (Data Science) в комп'ютерних
системах»**

**ОСНОВИ НАУКИ ПРО ДАНІ
Робоча програма навчальної дисципліни (Силабус)**

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Перший (бакалаврський)</i>
Галузь знань	<i>F(12) Інформаційні технології</i>
Спеціальність	<i>F2(121) Інженерія програмного забезпечення; F7(123) Комп'ютерна інженерія, F6(126) Інформаційні системи та технології</i>
Освітня програма	<i>Комп'ютерні системи та мережі Інженерія програмного забезпечення комп'ютерних систем Інтегровані інформаційні системи Інформаційні управляючі системи та технології Інформаційне забезпечення робототехнічних систем Інженерія програмного забезпечення інформаційних систем</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>Очна</i>
Рік підготовки, семестр	<i>3 курс, осінній семестр</i>
Обсяг дисципліни	<i>4 кредити/120 год. денна форма: лекцій 36 годин, Пр. роб. 18 год., СРС 66 год.</i>
Семестровий контроль/ контрольні заходи	<i>Залік, лабораторні роботи, МКР</i>
Розклад занять	<i>Згідно розкладу на 5-й семестр поточного навчального року за адресою https://schedule.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, https://bit.ly/3XAqgCq Лабораторні: д.т.н, професор, Новотарський Михайло Анатолійович novotariskyi.mykhailo@iit.kpi.ua</i>
Розміщення курсу	<i>Навчальний курс розміщений на платформі дистанційного навчання «Сікорський» в середовищі Google Workspace for Education: https://classroom.google.com/c/NDU4MDUzNDQ0MDU4?cjc=Iwqulmf</i>

Програма навчальної дисципліни

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

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

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 урахуванням 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. Розглядаються основні етапи аналізу даних та методи, які в цьому випадку використовуються. Детально розглянуто такі задачі добування даних, як класифікація та кластеризація, 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 очищення даних та аналізу отриманих результатів.

Предметом дисципліни є:

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

Студент буде 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.

Завданням дисципліни є формування і підсилення таких компетентностей:

Для спеціальності 121 Інженерія програмного забезпечення

- Здатність до абстрактного мислення, аналізу та синтезу.
- Здатність застосовувати знання у практичних ситуаціях.
- Здатність вчитися і оволодівати сучасними знаннями.
- Здатність ідентифікувати, класифікувати та формулювати вимоги до програмного забезпечення.
- Здатність формулювати та забезпечувати вимоги щодо якості програмного забезпечення у відповідності з вимогами замовника, технічним завданням та стандартами.
- Здатність накопичувати, обробляти та систематизувати професійні знання щодо створення і супроводження програмного забезпечення та визнання важливості навчання протягом всього життя.
- Здатність до алгоритмічного та логічного мислення.

Для спеціальності 123 «Комп'ютерна інженерія»

- Здатність вчитися і оволодівати сучасними знаннями.
- Здатність працювати в команді.
- Здатність використовувати сучасні методи і мови програмування для розроблення алгоритмічного та програмного забезпечення.
- Здатність проектувати системи та їхні компоненти з урахуванням усіх аспектів їх життєвого циклу та поставленої задачі, включаючи створення, налаштування, експлуатацію, технічне обслуговування та утилізацію.
- Здатність розробляти, адаптувати, використовувати програмно-технічні засоби для збирання, оброблення та інтелектуального аналізу даних в комп'ютерних системах: паралельних, розподілених, хмарних, безпечних, інтелектуальних, розумних, Інтернет речей, зокрема, з використанням штучного інтелекту.

Для спеціальності 126 Інформаційні системи та технології

- Здатність до абстрактного мислення, аналізу та синтезу.
- Здатність аналізувати об'єкт проектування або функціонування та його предметну область.
- Здатність застосовувати стандарти в області інформаційних систем та технологій при розробці функціональних профілів, побудові та інтеграції систем, продуктів, сервісів і елементів інфраструктури організації.
- Здатність до проектування, розробки, налагодження та вдосконалення системного, комунікаційного та програмно апаратного забезпечення інформаційних систем та технологій.
- Здатність управляти та користуватися сучасними інформаційно-комунікаційними системами та технологіями (у тому числі такими, що базуються на використанні Інтернет).
- Здатність розробляти структури робототехнічних систем та проектувати функціональні пристрої роботів та робототехнологічних комплексів.
- Здатність формувати нові конкурентоспроможні ідеї й реалізовувати їх у проектах (стартапах)

У відповідності до вищезазначеного, підсилені загальні та фахові компетенції дадуть наступні результати навчання:

- 1. Компетентність у розробці програмного забезпечення**, яке враховує особливості реалізації процесу KDD (Knowledge Discovery in Databases)
- 2. Вміння у повній мірі використовувати можливості бібліотеки PANDAS** для створення цільового набору даних
- 3. Здатність застосовувати структури даних Series та DataFrame** для аналізу одно- та багатовимірних наборів даних.
- 4. Здатність використовувати прийоми ефективного вибору метрик** для оцінки точності моделей.
- 5. Розробляти спеціалізоване програмне забезпечення** з використанням сучасних бібліотек для машинного навчання та аналізу даних (PyTorch, PyTensor).
- 6. Вміння відображати великі набори даних за допомогою бібліотеки Matplotlib.** Побудова діаграм та графіків
- 7. Підсилена здатність працювати з технічною документацією та дотримуватись стандартів проектування** – читати, створювати та застосовувати технічну документацію, дотримуватись норм ЕМС, вимог безпеки та виробничих стандартів у розробці електронних систем.

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

Навчальна дисципліна ґрунтується на попередніх знаннях , які охоплюють теми з курсів вищої математики: диференціальні обчислення, лінійна алгебра. Необхідні навички програмування алгоритмічною мовою Python, знання з комбінаторики, теорії графів та теорії множин.

Знання та навички, отримані під час вивчення дисципліни «Основи науки про дані», можуть бути застосовані під час вирішення навчальних та практичних завдань, що виникають під час:

- аналізу великих наборів даних з метою видобування корисних знань;
- розробки програмного забезпечення у сфері машинного навчання;
- впровадження технологій ШІ в сучасні проекти;
- розробка ПЗ для робототехнічних та автоматизованих систем;
- розробка ПЗ для аналізу зображень та розпізнавання образів;
- розробка адаптивних алгоритмів для підвищення надійності кіберфізичних систем.

Здобуті навички можуть бути використані під час виконання курсових і бакалаврських проектів, а також у практикоорієнтованих ініціативах, таких як стартапи та хакатони, пов'язані з розробкою програмних систем штучного інтелекту.

3. Зміст навчальної дисципліни

Вступ.

Розділ 1. Вступ у обробку великих наборів даних.

Лекція 1

Тема 1.1. Основні поняття про добування даних, процес KDD

Тема 1.2. Об'єктивні закономірності, принцип Бонферроні

Тема 1.3. Огляд методів та алгоритмів попередньої обробки даних

Розділ 2. Основні положення аналізу даних.

Лекція 2

Тема 2.1. Приклади задач аналізу даних

Тема 2.2. Поняття про міри та відстані у даних

Розділ 3. Базові поняття про бібліотеку Pandas.

Лекція 3

Тема 3.1. Архітектура бібліотеки Pandas

Тема 3.2. Тип даних Series основні операції з даним типом

Лекція 4

Тема 3.3. Створення модифікація та видалення наборів даних типу DataFrame

Тема 3.4. Арифметичні методи, які є застосовними до наборів даних типу DataFrame

Тема 3.5. Арифметичні операції між наборами даних типу Series та DataFrame

Розділ 4. Основні функції бібліотеки Pandas

Лекція 5

Тема 4.1. Застосування функцій pivot() та pivot_table() до повороту структур даних

Тема 4.2. Методи stacking та unstacking для перетворення наборів даних типу DataFrame

Тема 4.3. Застосування функції melt() до наборів даних типу DataFrame

Тема 4.4. Застосування функцій unfans для обробки елементів Series та DataFrame

Лекція 6

Тема 4.5. Функції для множин та функції статистики

Тема 4.6. Сортування для наборів даних типу Series та DataFrame

Тема 4.7. Кореляція та коваріантність для наборів даних типу Series та DataFrame

Тема 4.8. Інструменти вводу-виводу бібліотеки Pandas

Тема 4.9. Визначення пропущених даних та методи заповнення пропущених елементів

Тема 4.10. Злиття структур даних типу DataFrame

Розділ 5. Індексція та реіндексція наборів даних

Лекція 7

Тема 5.1. Вирівнювання даних у Pandas. Зміна порядку виводу наявних даних з маркером відсутності

Тема 5.2. Метод reindex() для Series та для DataFrame

Тема 5.3. Реіндексція з ключовим словом axis. Використання індексції стороннього набору даних.

Тема 5.4. Метод align(). Вирівнювання об'єктів з використанням align(s, join="outer"), align(s, join="inner"), align(s, join="left"), align(s, join="right").

Тема 5.5. Метод переіменування та метод переіменування з відображенням. Динамічне перейменування по осях.

Розділ 6. Багаторівнева індексція. Створення та використання об'єкта Multiindex.

Лекція 8

Тема 6.1. Методи створення об'єкта Multiindex

Тема 6.2. Керування відображенням індексів набору даних.

Тема 6.3. Розширене індексування з ієрархічним індексом з методом loc().

Тема 6.4. Використання зрізів. Перестановка рівнів за допомогою swaplevel.

Тема 6.5. Бінінг. Створення діапазонів інтервалів.

Тема 6.6. Методи агрегування даних

Розділ 7. Основи машинного навчання

Лекція 9.

Тема 7.1. Визначення машинного навчання та його класичні парадигми

- Тема 7.2. Глибинне навчання та його варіанти
- Тема 7.3. Статистичні та ймовірнісні методи
- Тема 7.4. Види вибірок (навчальна, тестова та контрольна)
- Тема 7.5. Об'єкти, ознаки та їх характеристики
- Тема 7.6. Принципи побудови регресійної моделі

Лекція 10.

- Тема 7.7. Бібліотека Scikit-learn та її застосування
- Тема 7.8. Моделі алгоритму у машинному навчанні
- Тема 7.9. Поняття про метод машинного навчання
- Тема 7.10. Функція втрат та функціонал якості
- Тема 7.11. Основні метрики у машинному навчанні.

Лекція 11

- Тема 7.12. Мінімізація емпіричного ризику
- Тема 7.13. Задача поліноміальної регресії у бібліотеці Scikit-learn
- Тема 7.14. Асимптотичний приклад мінімізації емпіричного ризику
- Тема 7.15. Спроможний метод. Емпіричні оцінки спроможності методу
- Тема 7.16. Метрики регресії.
- Тема 7.17. Методи класифікації

Лекція 12

- Тема 7.18. Метод k-найближчих сусідів.
- Тема 7.19. Нормалізація даних
- Тема 7.20. Оцінка точності класифікації
- Тема 7.21. Класифікація за допомогою методу k-найближчих сусідів
- Тема 7.22. Метод wKNN

Розділ 8. Основи PyTorch. Тензори

Лекція 13

- Тема 8.1. Способи створення тензора
- Тема 8.2. Перетворення Series та DataFrame у PyTorch tensor
- Тема 8.3. Перетворення типів даних PyTorch
- Тема 8.4. Модифікація тензора
- Тема 8.5. Арифметичні операції з тензорами
- Тема 8.6. Фільтрація і конкатенація тензорів
- Тема 8.7. Тензорні згортки
- Тема 8.8. Стандартні функції для тензорів
- Тема 8.9. Зміна форми тензора

Розділ 9. Основи PyTorch. Задача логістичної регресії

Лекція 14

- Тема 9.1. Набір даних MNIST
- Тема 9.2. Завантаження бібліотек та початкові установки
- Тема 9.3. Перетворення формату рисунка
- Тема 9.4. Теоретичні основи логістичної регресії
- Тема 9.5. Обчислення крос-ентропії
- Тема 9.6. Крос-ентропійна функція втрат

Розділ 10. Бібліотека PyTensor

Лекція 15

- Тема 10.1. Завантаження бібліотеки PyTensor
- Тема 10.2. Основні принципи функціонування PyTensor

Тема 10.3. Створення функції з використанням Op
Тема 10.4. Операція над двома матрицями
Тема 10.5. Обчислення градієнтів
Тема 10.6. Обчислення Якобіану
Тема 10.7. Обчислення Гаусіана
Тема 10.8. Цикли, сканування

Розділ 11. Бібліотека TensorFlow

Лекція 16

Тема 11.1. Установка бібліотеки TensorFlow
Тема 11.2. Створення тензорів
Тема 11.3. Математичні операції над тензорами
Тема 11.4. Модуль Keras: Високорівневий API
Тема 11.5. Створення простої моделі
Тема 11.6. Компіляція моделі
Тема 11.7. Метрики у задачах TensorFlow

Розділ 12. Відображення даних

Лекція 17

Тема 12.1. Встановлення бібліотеки Matplotlib
Тема 12.2. Відображення графіків з використанням бібліотеки Matplotlib
Тема 12.3. Побудова гістограм та діаграм
Тема 12.4. Криві Ендрюса
Тема 12.5. Стили відображення графіків

Лекція 18

Тема 12.6. Встановлення бібліотеки Seaborn
Тема 12.7. Основні характеристики бібліотеки Seaborn
Тема 12.8. Огляд додаткових тем Seaborn
Тема 12.9. Застосування високорівневого API
Тема 12.10. Застосування кумулятивних функцій
Тема 12.11. Багатовимірні перегляди складних наборів даних
Тема 12.12. Використання довгого та широкого форматів

4. Навчальні матеріали і ресурси

4.1. Базова література:

1. Основи науки про дані [Електронний ресурс]: навч. посіб. для студ. спеціальності 123,121 / М. А. Новотарський; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,5 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2022. – 294 с.
2. Новотарський М.А. Лекції х курсу «Основи науки продані» // <https://cloud.comsys.kpi.ua/s/pjw4nAStZFPsDos>
3. Новотарський М.А. Методичні вказівки до виконання лабораторних робіт з курсу «Основи науки продані» // Новотарський М.А. Лекції з курсу «Основи науки продані» // <https://classroom.google.com/c/NDU4MDUzNDQ0MDU4?cjc=lwqulmf>
4. Системи Data Science. Практикум з системної інженерії систем Data Science [Електронний ресурс] / КПІ ім. Ігоря Сікорського ; уклад. П. П. Масляноко. – Електронні текстові дані (1 файл: 2,3 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2022. – 68 с.
5. Новотарський М.А., Нестеренко Б.Б. Штучні нейронні мережі: обчислення.– К.: Інститут математики НАН України, 2004.– 408 с.

6. Нестеренко Б.Б., Новотарський М.А. Методи розв'язування крайових задач на нейронних мережах // Збірник праць Інституту математики НАН України.– 2006.– Т.3, №4.– С.231-238.

7. Нестеренко Б.Б., Новотарський М.А. Формальні засоби моделювання паралельних процесів та систем // Праці Інституту математики НАН України.– Т.90.–Київ: Ін-т математики НАН України, 2012.–334с.

4.2. Додаткова література:

1. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. – К.: ІСЗІ КПІ ім. Ігоря Сікорського», 2018. — 297с.
2. Гороховатський, Володимир Олексійович. Методи інтелектуального аналізу та оброблення даних: навчальний посібник / В.О. Гороховатський, І.С. Творошенко ; Міністерство освіти і науки України, Харківський національний університет радіоелектроніки. – Харків : ХНУРЕ, 2021. – 90 с. https://opac.kpi.ua/F/?func=direct&doc_number=000634401&local_base=KPI01
3. Інтелектуальний аналіз даних : навчальний посібник / укладачі: Я.М. Дрінь, І.В. Малик, Ю.А. Літвінчук ; Міністерство освіти і науки України, Чернівецький національний університет імені Юрія Федьковича. – Чернівці : Чернівецький національний університет імені Юрія Федьковича, 2020. – 95 с. https://opac.kpi.ua/F/?func=direct&doc_number=000615829&local_base=KPI01
4. Юрченко, Ігор Валерійович. Прикладний статистичний аналіз з використанням Python : навчальний посібник / І.В. Юрченко ; Міністерство освіти і науки України, Чернівецький національний університет імені Юрія Федьковича. Видання 3-є, доповнене. – Чернівці : Чернівецький національний університет, 2021. – 107 стор. Замовити в Бібліотеці КПІ: https://opac.kpi.ua/F/?func=direct&doc_number=000635029&local_base=KPI01

4.1. Інформаційні ресурси

1. Основи науки про дані / дистанційний курс // <https://classroom.google.com/c/NDU4MDUzNDQ0MDU4?cjc=lwqulmf>
2. <https://www.kaggle.com/>
3. <https://www.jetbrains.com/>
4. <https://github.com/PacktPublishing/Artificial-Intelligence-with-Python>
5. <https://pandas.pydata.org/>
6. <https://scapy.net/>
7. <https://developers.google.com/optimization>
8. <https://www.tensorflow.org/>
9. <https://keras.io/>
10. <https://opencv.org/>
11. <https://numpy.org/doc/stable/reference/generated/numpy.polyfit.html>
12. <https://www.statsmodels.org/stable/examples/notebooks/generated/ols.html>
13. <https://scikit-learn.org/stable/modules/sgd.html#regression>

5. Методика опанування навчальної дисципліни (освітнього компонента)

Розподіл навчального часу за видами занять і завдань з дисципліни згідно з робочим навчальним планом: на кредитний модуль виділено 120 годин та 4 кредитів – 18 лекцій (36 годин), 5 лабораторних робіт (18 годин), самостійна робота (66 годин), заліковий тест.

Лекційні заняття

Для досягнення мети навчальної дисципліни **в лекційному матеріалі** студенти мають отримати знання про можливі види даних та способи їх обробки, про типові задачі, які виникають при обробці даних. Особлива увага приділяється використанню сучасних інструментів попередньої підготовки даних, зокрема бібліотеки Pandas, бібліотеки Matplotlib та ряду інших бібліотек, використання яких дозволяє пройти повний цикл підготовки даних перед обробкою. Це забезпечить студентам розуміння важливості попереднього аналізу даних та знання технологій підготовки цільових наборів даних у рамках процесу KDD. У рамках курсу також передбачено знайомство з основними принципами аналізу даних на базі бібліотек TensorFlow і PyTorch.

Деталізація тем лекційних занять, конспект лекцій та відеоматеріал розміщено на платформі дистанційного навчання «Сікорський» за посиланням :

<https://classroom.google.com/c/NDU4MDUzNDQ0MDU4?cjc=lwqulmf>

Лабораторний практикум.

Дані лабораторні роботи виконують шляхом модифікації початкових даних, які знаходяться у зошиті, номер якого відповідає номеру лабораторної роботи. Тому необхідно попередньо створити папку та завантажити в цю папку всі початкові зошити. Необхідно також уміти створити оригінальні зошити jupyter notebook.

Для успішного захисту лабораторних робіт разом зі звітами викладачеві необхідно надати зошити з виконаними завданнями для кожної лабораторної роботи. Звіти та зошити jupyter notebook необхідно завантажувати у відповідне завдання, яке створене системі Google classroom. Перевірка теоретичних знань проводиться персонально виключно на заняттях з лабораторних робіт у режимі відеоконференції.

Мета лабораторних робіт — набуття практичних навичок у розробці програмних застосунків для початкової обробки даних та формування цільових наборів даних, а саме:

- Технологія вибору відповідних даних з DataFrame або Series, вибір фрагментів даних з набору даних.
- Технологія швидкого та ефективного вибору необхідних даних з набору даних з використанням структур даних та інструментів бібліотеки Pandas.
- Технологія роботи з наборами даних, які містять відсутні або помилкові дані.
- Робота з простою моделлю лінійної регресії.
- Робота з простим алгоритмом кластеризації, який відносять до алгоритмів аналізу ринкового кошику.

Тематика лабораторних робіт

Лабораторна робота 1 . Підсумкові функції та відображення.

Мета: отримати початкові знання про середовище Jupyter notebook та навчитися створювати зошити.

Лабораторна робота 2. Практичне застосування бібліотек NumPy, SciPy та Matplotlib

Мета: вивчити основний функціонал та навчитися практично застосовувати бібліотеки NumPy, SciPy та Matplotlib до вирішення інженерних задач.

Лабораторна робота 3. Індексвання, вибір, редагування набору даних.

Мета: навчитися швидко та ефективно вибирати необхідні дані з набору даних з використанням структур даних та інструментів бібліотеки Pandas.

Лабораторна робота 4. Підсумкові функції та відображення.

Мета: навчитися вибирати відповідні дані з DataFrame або Series, а також правильно вибирати дані з набору даних.

Лабораторна робота 5. Робота з пропущеними даними.

Мета: навчитися працювати з наборами даних, які містять відсутні або помилкові дані.

6. Самостійна робота здобувача вищої освіти денної форми навчання

Види самостійної роботи для студентів денної форми навчання (**66 годин**):

- підготовка до аудиторних занять, виконання поточних домашніх завдань та опрацювання матеріалів лекцій (**0,5 годин x 36 годин = 18 годин**);
- виконання індивідуального завдання до лабораторних робіт, підготовка та оброблення проведення розрахунків, оформлення звіту до лабораторної роботи (1 година x 18 годин лабораторних занять (5 лабораторних робіт) = **18 годин**);
- підготовка до заліку (**2 годин**);
- самостійне опрацювання теоретичного матеріалу (**20 годин**);
- розгортання програмного середовища та завантаження вихідних кодів програмного забезпечення для виконання лабораторних робіт (**80 годин**).

Тематика для самостійного опрацювання теоретичного матеріалу наведена в таблиці 1.

Таблиця 1

Тематика самостійної роботи студента

№ з/п	Назва теми лекції та перелік основних питань (перелік дидактичних засобів, посилання на літературу та завдання на СРС)	Кількість годин на СРС
1	Вступ. Базові відомості про дані. Види даних та способи їх обробки. Операції над даними. Класифікація інформації за видами і типами. Види кодувань. Структура даних та одиниці вимірювання. Література: https://hi-news.pp.ua/kompyuteri/1379-vidi-danih-sposobi-yih-obrobki.html	2
2	Основні положення аналізу даних. Поняття про взаємозв'язок статистичних показників. Види та форми зв'язків між явищами. Прийоми виявлення щільності зв'язку між показниками досліджуваних явищ. Література: https://xreferat.com/22/2260-1-osnovn-polozhennya-kompleksnogo-statistichnogo-anal-zu-danih-u-pravov-iy-statistic.html	2
3	Базові поняття про бібліотеку Pandas. Загальні поняття про Pandas. Сфера застосувань бібліотеки Pandas. Додаткові властивості структур даних у Pandas/ Література: https://coderlessons.com/tutorials/mashinnoe-obuchenie/uchebnik-tensorflow/6-uchebnik-po-python-pandas	2
4	Основні характеристики та способи застосування структури даних DataFrame. class pandas.DataFrame. Атрибути та методи структури даних DataFrame Література: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html	2
5	Функції бібліотеки Pandas. Модифікація великих наборів даних. Бінарні функції. Функції порівняння значень. Перевірка на еквівалентність. Порівняння масивів. Об'єднання даних, які перекриваються. Методи описової статистики. Підрахунок значень та мода значень. Дискретизація неперервних значень. Література: https://pyprog.pro/pd/pd_11_basic_functionality.html	2

№ з/п	Назва теми лекції та перелік основних питань (перелік дидактичних засобів, посилання на літературу та завдання на СРС)	Кількість годин на СРС
6	Індексація та реіндексація наборів даних. Різні варіанти індексації. Методи для індексації. Література: https://coderlessons.com/tutorials/python-technologies/vyuchit-python-panda/python-pandas-indeksirovanie-i-vybor-dannykh	2
7	Багаторівнева індексація. Створення та використання об'єкта Multiindex. 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. Література https://pandas.pydata.org/docs/user_guide/advanced.html	2
8	Відображення великих наборів даних. Бібліотека Matplotlib. Оформлення графіків. Розташування графіків. Робота з осями. Інтерактивні можливості. Література: https://jenyay.net/Matplotlib/Matplotlib	2
9	Задача логістичної регресії з використанням бібліотеки PyTorch. Установка PyTorch. Підготовка даних в PyTorch. Навчання моделі за допомогою PyTorch. Розгортання моделі за допомогою Windows. Література: http://cloud-5.bitp.kiev.ua/?page_id=605 https://docs.microsoft.com/ru-ru/windows/ai/windows-ml/tutorials/pytorch-installation	2
10	Створення глибокої нейронної мережі з використанням PyTorch. Класифікація зображень за допомогою PyTorch. Проблема класифікації. Завантаження даних PyTorch. Створення навчального набору даних. Волідація та контрольні набори даних. Література: https://habr.com/ru/company/piter/blog/512158/	2

Всього 20 годин

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Дедлайни. Для виконання практичних робіт встановлюються дедлайни. Дедлайни уточнюються і оприлюднюються викладачем на початку семестру.

Заохочувальні бали (бонуси) виставляються за: активну участь на лекціях, виконання додаткової лабораторної роботи. Кількість заохочуваних балів встановлює викладач, але не більше 10.

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

В процесі навчання викладач має право нарахувати заохочувальний бал 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.

Правила відвідування занять. Відвідування занять є вільним, бали за присутність на лекціях та практичних заняттях не додаються, відповідно штрафні бали не передбачаються. Втім, вагома частина рейтингу формується через активну участь у заходах на практичних заняттях.

Дистанційний режим навчання. У разі запровадження обмежень на відвідування університету, пов'язаних з введенням карантину або режиму воєнного стану в державі, освітній процес здійснюється у дистанційному режимі відповідно до Положення про дистанційне навчання в КПІ ім.

Ігоря Сікорського (<https://osvita.kpi.ua/index.php/node/188>), Регламенту організації освітнього процесу в дистанційному режимі (<https://profkom.kpi.ua/reglament-organizatsiyi-osvitnogo-protsesu-v-distantsiynomu-rezhimi>) та Регламенту проведення семестрового контролю в дистанційному режимі (<https://osvita.kpi.ua/node/148>). У режимі дистанційного навчання заняття відбуваються у вигляді онлайн-конференції на платформах BigBlueButton, Google Meet, Zoom. Посилання на конференцію видається на початку семестру і розміщується в АС «Електронний кампус». З метою забезпечення якісної підготовки здобувачів, дистанційний курс дисципліни розміщено на Платформа дистанційного навчання «Сікорський» (<https://www.sikorsky-distance.org>). Результати оцінювання висвітлюють у АС «Електронний кампус» на особистій сторінці здобувача (<https://ecampus.kpi.ua>).

Правила поведінки на заняттях. На заняттях слід дотримуватись норм етичної поведінки визначених у Кодексі честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>). На території університету здобувачі мають поводити себе відповідно до Правил внутрішнього розпорядку (<https://kpi.ua/admin-rule>).

Визнання результатів навчання, набутих у неформальній/інформальній освіті. Положення про визнання результатів навчання, набутих у неформальній / інформальній освіті (<https://osvita.kpi.ua/index.php/node/179>) регламентує визнання результатів навчання, набутих у неформальній/інформальній освіті:

- у разі навчання на неформальній освіті, яку студент обрав самостійно і хоче зарахувати результати навчання в рамках дисципліни, проходиться процедура валідації, що передбачає подання здобувачем заяви на ім'я декана, декларації підтверджувальних документів; рішення про визнання чи не визнання приймається комісією у складі завідувача кафедри, викладача, гаранта освітньо-професійної програми.

- за рішенням викладача окремі змістовні модулі або теми дисципліни можуть бути зараховані на підставі результатів проходження здобувачем онлайн-курсу або іншого елементу неформальної освіти; перелік завдань та елементів неформальної освіти визначається викладачем заздалегідь і розміщується на платформі дистанційного навчання «Сікорський», про що студенти повідомляються на початку семестру; у цьому випадку здобувач звільняється від виконання відповідних завдань і отримує максимальний бал згідно з рейтинговою системою оцінювання.

Політика використання штучного інтелекту. Використання штучного інтелекту регламентується «Політикою використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського» (<https://osvita.kpi.ua/node/1225>). Усі завдання, як під час виконання навчальних завдань з дисципліни, так і індивідуальні завдання, мають бути результатом власної оригінальної роботи здобувача. Використання ШІ для автоматичної генерації відповідей без подальшого їх аналізу та доопрацювання заборонено.

Порушення академічної доброчесності: використання плагіату. У разі виявлення фактів використання плагіату під час виконання навчальних завдань, студент втрачає право на подальше виконання поточних завдань у межах навчальної дисципліни до з'ясування обставин порушення.

Студент зобов'язаний аргументовано пояснити причини використання плагіату, а також засвідчити готовність дотримуватись принципів академічної доброчесності в подальшому навчанні (**письмово у вигляді пояснювальної записки**); повторно виконати завдання або, за потреби, виконати додаткове завдання відповідно до бачення викладача з урахуванням вимог щодо самостійності.

Зазначені дії регламентуються Кодексом честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>).

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

Види контролю з навчальної дисципліни «Основи науки про дані» включають:

Лабораторні роботи:

Заплановано самостійне виконання 5 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 технологій початкової обробки даних.

Поточний контроль:

Передбачено 2 календарні тестування закритими тестами у системі TCEXAM, які повністю охоплюють тематику даної навчальної дисципліни. Кожний поточний закритий тест містить 12 питань та триває 15 хв. 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.

Залік

Необхідною умовою допуску студента до заліку є його індивідуальний семестровий рейтинг 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.

Студенти можуть зараховувати виконання лабораторних робіт з даного курсу за рахунок отримання сертифікатів на навчальних платформах. Умови зарахування та кількість сертифікатів узгоджується з викладачем, але не більше, ніж 2 лабораторні роботи за один сертифікат.

Залік проводиться у вигляді співбесіди зі студентом для об'єктивного визначення рівня знань, умінь та практичних навичок, отриманих за семестр. За вибором викладача залік може проводитися у вигляді контрольного тесту у системі TCEXAM та складається з 20 питань. Правильна відповідь на кожне запитання тесту додає 1 бал.

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 2.

Таблиця 2

Оцінювання окремих видів навчальної роботи студента

Назва заняття	Форма контролю	Кількість балів	Допуск до семестрового контролю	Всього балів
Лабораторна робота №1	Звіт	4	6	10
	Опитування	3		
	Сирцевий код	3		
Лабораторна робота №2	Звіт	4	6	10
	Опитування	3		
	Сирцевий код	3		
Лабораторна робота №3	Звіт	4	6	10
	Опитування	3		
	Сирцевий код	3		
Лабораторна робота № 4	Звіт	4	6	10
	Опитування	3		
	Сирцевий код	3		
Лабораторна робота № 5	Звіт	4	6	

	Опитування	3		10
	Сирцевий код	3		
Календарний контроль 1		15	10	15
Календарний контроль 2		15	10	15
Залік		20	10	20
Усього за семестр		100	60	100

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

Календарний контроль: тематичні тестування закритими тестами –
2x15=30 балів.

Залік за умови тестування закритими тестами – 20 балів,
Виконання лабораторних робіт.

Протягом семестру студенти виконують 5 лабораторних робіт.

Максимальна кількість балів за кожну лабораторну роботу – 5+5=10 балів.

Разом за лабораторні роботи (максимальна кількість балів) – 50.

Виконання залікового контрольного тестування.

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.

Розрахунок розміру шкали (R) рейтингу:

Сума вагових балів контрольних заходів протягом семестру становить:

$R = \sum_k r_k$, де r_k максимальний рейтинговий бал кожного з контрольних заходів (контрольне тестування, лабораторні роботи, преміальні бали, мінус штрафні бали).

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 та залік).

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. 5).

Таблиця 5

Відповідність рейтингових балів оцінкам за університетською шкалою

Кількість балів	Оцінка
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

Порядок ліквідації студентами академічної заборгованості

Для студентів, що не виконали вчасно індивідуальний навчальний план деканатом призначається додаткова сесія і надається можливість двох перескладань. Час першого та другого перескладань

додатково узгоджується. Для студентів, яким призначена додаткова сесія поточний рейтинг анулюється і призначаються такі умови:

Для отримання позитивного перескладання студент має:

1. виконати всі лабораторні роботи,
2. пройти семестровий контроль і надати більше 60% правильних відповідей.

Залікова оцінка, яка отримана у додаткову сесію, не може перевищувати «достатньо»

9. Додаткова інформація з дисципліни (освітнього компонента)

Викладання дисципліни «Основи науки про дані» для спеціальностей 121, 123 і 126 має свою специфіку, яка пов'язана з тим, що аналіз даних потребує попередньої підготовки цільових наборів даних. Для формування таких наборів використовують процес KDD, який розділяє етапи попередньої підготовки даних. В рамках даних етапів використовують спеціальні бібліотеки, технології застосування яких описуються у даному навчальному курсі.

Робочу програму навчальної дисципліни (силабус):

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

Ухвалено кафедрою ОТ (протокол №18 від 24.06.2026)

Погоджено: Методичною комісією ФІОТ (протокол № 12 від 26.06.2026)