

ПРОГРАМА НАВЧАННЯ

2026–2027

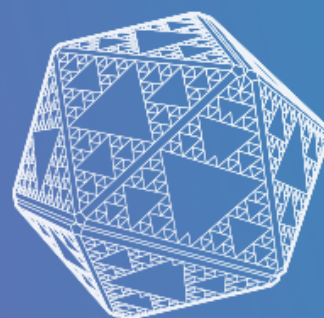
Data Science / AI / Big Data

B I G

D A T A



L A B



Зміст

ПІДГОТОВЧИЙ КУРС

Основи математики для аналітиків	4
Python для початківців	6

ОСНОВНИЙ КУРС

Основи лінійної алгебри, методів оптимізації, теорії ймовірностей та математичної статистики для науки про дані	8
Основи математичного аналізу для науки про дані	11
Зберігання даних, їх аналіз та візуалізація	13
Основи програмування та інженерії	15
Машинне навчання та інтелектуальний аналіз. Прогнозний аналіз і часові ряди	16
Python для машинного навчання	19
Нейронні мережі та Глибоке навчання	21
Інфраструктура та технології Big Data	22
Практичний модуль курсу з інфраструктури: «Від локального data product до хмари»	25
Сучасні технології штучного інтелекту та їх практичне застосування	26
DevOps Workshops	30

ПІДГОТОВЧИЙ КУРС

Основи математики для аналітиків

Math Basics for Analytics

Лектор	Ставицький Андрій Володимирович, доктор економічних наук, професор.
Тривалість курсу	24 аудиторні години 4 лекційні заняття 11 практичних занять 1 підсумкова контрольна робота Одне заняття - 1,5 години

ПРОГРАМА КУРСУ

Заняття 1 (лекція). Вступ.

- Задача дисципліни
- Організаційні задачі

Заняття 2 (лекція). Основні математичні перетворення.

- Розрахункові задачі
- Алгебраїчні перетворення
- Тригонометричні перетворення
- Логарифмічні перетворення
- Розв'язання рівнянь
- Розв'язання систем рівнянь

Заняття 3 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P1)
- Самостійна робота

Заняття 4 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P2)
- Самостійна робота

Заняття 5 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P3)
- Самостійна робота

Заняття 6 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P3)

Заняття 7 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P6)
- Самостійна робота

Заняття 8 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P6)
- Самостійна робота

Заняття 9 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P7)
- Самостійна робота

Заняття 10 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P8)
- Самостійна робота

Заняття 11 (лекція). Похідна.

- Визначення похідної
- Застосування похідної

Заняття 12 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P15)
- Самостійна робота

Заняття 13 (лекція). Інтеграли.

- Невизначені інтеграли
- Визначені інтеграли
- Способи інтегрування

Заняття 14 (практичне заняття). Розв'язання задач.

- Розв'язання практичних завдань (P15)
- Самостійна робота

Заняття 15 (Підсумкова контрольна робота).

- Розв'язання задач

Заняття 16 (практичне заняття). Розв'язання задач.

- Розбір підсумкової роботи
- Самостійна робота

Python для початківців

Python for beginners

Лектор	Сергій Одрібець, к. ф.-м. н., Senior Data Scientist, Vodafone Ukraine
Тривалість курсу	24 аудиторні години 2 - лекційних заняття 6 - лекційно-практичних занять

ПРОГРАМА КУРСУ

Модуль 1.

Введення в Python: переваги та недоліки мови, середовище розробки, встановлення та налаштування, бібліотеки та менеджер пакетів, синтаксис мови та основні типи даних, динамічна типізація, ключові слова та вбудовані функції, перша програма. Самостійна робота.

Модуль 2.

Числові типи даних та операції над числами. Перетворення чисел. Логічний тип boolean та логічні операції. Самостійна робота.

Модуль 3.

Умовні оператори та цикли. Інструкції break, continue, pass. Обробка виключень. Самостійна робота.

Модуль 4.

Список (List), індекси та зрізи. Методи, функції та операції зі списками. Генератори та ітератори. Самостійна робота.

Модуль 5.

Рядок (String), функції та методи. Екрановані послідовності та форматування. Множина (Set) та операції над множинами. Самостійна робота.

Модуль 6.

Кортеж (Tuple), функції та методи. Словник (Dictionary), функції та методи. Самостійна робота.

Модуль 7.

Функції. Область видимості змінних. Знайомство з бібліотеками візуалізації matplotlib, seaborn, plotly. Самостійна робота.

Модуль 8.

Основи ООП: класи, конструктори, успадкування, інкапсуляція, поліморфізм, декоратори.

ОСНОВНИЙ КУРС

Основи лінійної алгебри, методів оптимізації, теорії ймовірностей та математичної статистики для науки про дані

Linear Algebra, Optimization Methods, Probability Theory and Mathematical Statistics fundamentals for Data Science

Лектор	Ставицький Андрій Володимирович, доктор економічних наук, професор.
Тривалість курсу	24 аудиторні години 7 лекційних занять 8 практичних занять 1 екзаменаційна контрольна робота Тривалість кожного заняття - 1,5 години.

ПРОГРАМА КУРСУ

Заняття 1 (лекція).

Вступ до лінійної алгебри.

- Вектори
- Матриці
- Дії з матрицями
- Знаходження визначника матриці

Заняття 2 (практичне заняття).

Робота з матрицями.

- Розв'язання задач

Заняття 3 (лекція).

Спеціальні типи матриць.

- Власні значення та власні вектори
- Симетричні матриці та ортогональна діагоналізація
- Застосування власних значень та власних векторів
- Аналіз основних компонентів
- Складні дії з матрицями

Заняття 4 (практичне заняття).

Визначення власних значень та векторів.

- Застосування власних значень та власних векторів
- Розв'язання практичних завдань
- Самостійна робота

Заняття 5 (лекція).

Огляд методів оптимізації.

- Методи одновимірної мінімізації

- Методи багатовимірної мінімізації
- Основні завдання динамічного програмування

Заняття 6 (практичне заняття).

Розв'язання типових задач оптимізації.

- Розв'язання практичних завдань
- Самостійна робота

Заняття 7 (практичне заняття).

Огляд методів комбінаторики.

- Факторіал
- Комбінації, перестановки, розміщення
- Розв'язання практичних завдань
- Самостійна робота

Заняття 8 (лекція).

Теорія ймовірностей.

- Робота з множинами
- Ймовірність та різновиди її знаходження
- Формула умовної ймовірності
- Формула повної ймовірності
- Формула Байєса

Заняття 9 (практичне заняття).

Визначення ймовірності на практиці.

- Робота з простором подій
- Розв'язання практичних завдань визначення ймовірності подій
- Самостійна робота

Заняття 10 (лекція).

Розподіли випадкових величин.

- Випадкова величина
- Математичне сподівання та дисперсія
- Формула Бернуллі та біноміальний розподіл
- Розподіл Пуассона
- Нормальний розподіл

Заняття 11 (практичне заняття).

Застосування розподілів випадкових величин на практиці.

- Розв'язання практичних завдань

Заняття 12 (лекція).

Описова статистика.

- Вибіркове середнє, медіана
- Вибіркова дисперсія
- Інтервальні та дискретні вибірки

Заняття 13 (практичне заняття).

Математична статистика.

- Визначення параметрів випадкової величини
- Побудова гістограми, емпіричної функції розподілу
- Самостійна робота

Заняття 14 (лекція).

Перевірка статистичних гіпотез.

- Види гіпотез та етапи їх перевірки
- Правила перевірки гіпотез

Заняття 15 (практичне заняття).

Перевірка статистичних гіпотез.

- Гіпотеза про значення вибіркового середнього вибірки
- Гіпотеза про різницю вибірових середніх вибірок
- Гіпотеза про значення дисперсії вибірки
- Гіпотеза про рівність дисперсій вибірок

Заняття 16 (Підсумкова контрольна робота).

Іспит.

- Розв'язання задач

Основи математичного аналізу для науки про дані

Calculus fundamentals for Data Science

Лектор	Чертов Олег Романович, доктор технічних наук, професор
Тривалість курсу	12 аудиторних годин 1 лекційне заняття 6 практичних занять 1 екзаменаційна контрольна робота Тривалість кожного заняття: 1,5 години.

ПРОГРАМА КУРСУ

Заняття 1 (лекція).

Вступ.

- Строгість, точність та абстрактність математики.
- Логічна символіка (висловлення, логічні зв'язки, квантори).
- Множини, операції над ними.
- Бінарні відношення. Відношення еквівалентності. Клас еквівалентності.
- Відношення порядку. Обмежена множина. Супремум та інфімум. Максимум та мінімум.
- Відображення (функція). Бієкція. Обернене відображення. Композиція функцій. Послідовність.

Заняття 2 (практичне заняття).

Математична індукція. Границі послідовностей.

- Метод математичної індукції, його застосування до розв'язання задач.
- Означення границі послідовності, її геометрична інтерпретація.
- Пошук границь послідовностей.

Заняття 3 (практичне заняття).

Границі функцій.

- Означення границі функції, її геометрична інтерпретація.
- Пошук границь функцій.
- Неперервні функції. Точки розриву функцій.

Заняття 4 (практичне заняття).

Похідна. Дослідження функцій.

- Означення похідної функції однієї змінної, її геометрична інтерпретація.
- Пошук похідних та диференціалів (у тому числі вищих порядків) функцій однієї змінної, функцій заданих параметрично і заданих неявно, складеної та оберненої функції.
- Дослідження функцій однієї змінної на екстремуми та опуклість, побудова графіків функцій однієї змінної.

Заняття 5 (практичне заняття).

Невизначені та визначені інтеграли.

- Властивості невизначених інтегралів. Внесення під знак диференціалу. Інтегрування частинами.
- Знаходження первісних раціональних, ірраціональних та трансцендентних функцій.
- Обчислення визначених інтегралів від функції однієї змінної, їх застосування до обчислення довжин кривих і площ фігур.

Заняття 6 (практичне заняття).

Ряди Тейлора та Маклорена. Числові та функціональні ряди.

- Побудова поліномів Тейлора і Маклорена для функцій однієї змінної і використання їх до наближення функцій.
- Поточкова та рівномірна збіжність рядів. Дослідження на збіжність числових та функціональних рядів.
- Знаходження області та радіусу збіжності степеневих рядів.

Заняття 7 (практичне заняття).

Частинні похідні та диференціал функції багатьох змінних. Градієнт.

- Знаходження частинних похідних та диференціалів функції багатьох змінних, градієнту функції.
- Поняття локального та умовного екстремуму функцій багатьох змінних.

Заняття 8 (екзамен).

Підсумкова контрольна робота.

- Шість задач/вправ по темам, що були засвоєні/повторені під час курсу.

Зберігання даних, їх аналіз та візуалізація

Data Warehousing, Business Intelligence and Data Visualization

Лектор	Рижук Володимир Володимирович, Senior Data Analyst / Data Scientist, Nova Poshta
Тривалість курсу	24 аудиторні години 8 занять по 3 години

ПРОГРАМА КУРСУ

Заняття 1.

Дані в бізнесі та «погані дані»

- роль даних в організації, типи даних і метадані, SQL vs NoSQL, огляд хмарних сервісів;
- огляд курсу, наскрізного кейсу ProZorro та фінального продукту (демо PWA);
- практика: аудит якості сирих даних ProZorro (JSON), запуск навчального середовища (PostgreSQL у Docker, DBeaver, GitHub, Tableau Public);
- формування груп і вибір теми групового проєкту.

Заняття 2.

Моделювання даних

- концептуальне, логічне та фізичне проектування; ER-модель, ключі та зв'язки;
- нормалізація до третьої нормальної форми (3NF);
- практика: перетворення JSON закупівлі на ER-модель (замовник, лот, предмет, пропозиція, договір), створення таблиць у PostgreSQL та завантаження шару raw скриптом.

Заняття 3.

SQL для аналітики. Частина 1

- мови SQL (DDL, DML, DCL), підключення до бази через DBeaver;
- SELECT, WHERE, JOIN, GROUP BY, HAVING;
- практика: відповіді на бізнес-питання до даних ProZorro з автоматичною перевіркою результатів.

Заняття 4.

Сховище даних і вимірне моделювання

- DB vs DWH; факти, виміри, міри, зерно (grain); star і snowflake схеми; SCD;
- огляд архітектур: data mart, data lake, lakehouse, data mesh; словник даних;
- практика: star schema для закупівель пального (факт «договір», виміри замовник, постачальник, CPV, дата, регіон);
- чекпоінт 1 групового проєкту: модель даних і перелік дефектів.

Заняття 5.

SQL для аналітики. Частина 2 та якість даних

- CTE, віконні функції, SQL-views та процедури;
- Master Data Management і golden record, життєвий цикл даних;
- практика: дедуплікація постачальників за ЄДРПОУ, нормалізація одиниць виміру та назв пального, перерахунок валют за курсами НБУ, вітрини mart_*.

Заняття 6.

Конвеєри даних (ETL/ELT), Data Governance та прототип PWA

- ETL vs ELT, пакетні та потокові конвеєри, шари raw → staging → mart;
- Data Governance: ролі та права доступу (GRANT), окрема роль для читання вітрин;
- практика: ETL-скрипт і процедура оновлення вітрин; збірка проєкту в docker compose; прототип PWA за допомогою AI-інструментів;
- чекпоінт 2 групового проєкту: проєкт піднімається з нуля однією командою, PWA працює локально.

Заняття 7.

Бізнес-аналітика в Tableau

- що таке BI, OLAP vs OLTP, типи звітів, дерево KPI;
- практика: вивантаження вітрин, побудова KPI-дашборду в Tableau Public, публікація та вбудовування в PWA.

Заняття 8.

Візуалізація даних, data storytelling та захист проєктів

- найкращі практики та поширені помилки у візуалізації, структура data story;
- практика: критика дашбордів;
- захист групових проєктів (до 10 хвилин на групу), зворотний зв'язок, підсумки курсу.

Основи програмування та інженерії

Programming and Engineering Basics

Лектор	Сергій Одрібець. Data scientist, Vodafone Ukraine
Тривалість курсу	21 аудиторна година
	5 - лекційних занять
	7 - практичних занять
	2 - лекційно практичних заняття
	1 - контрольна робота/код рев'ю

ПРОГРАМА КУРСУ

Заняття 1 (лекція).

- Введення в Python. Встановлення Python. IDE для Python. Встановлення бібліотек. Python virtual Environments. Написання першої програми

Заняття 2 (практичне заняття)

- Типи та структури даних: Числові, текстові. Словники, списки та кортежі в Python

Заняття 3 (лекція)

- Оператори та вирази в Python: If, elif, for, while, break, continue, else.

Заняття 4 (практичне заняття)

- Функції в Python. Лямбда функції. Рекурсія. Декоратори функцій

Заняття 5 (практичне заняття)

- Модулі та пакети в Python. Генератор та ітератори

Заняття 6 (лекція).

- Введення в концепції ООР. Побудова класів в Python. Абстракція, інкапсуляція

Заняття 6 (практичне заняття)

- ООР. Наслідування, поліморфізм в Python. Шаблони проектування. Виключення в Python

Заняття 7 (практичне заняття)

- Модуль Numpy

Машинне навчання та інтелектуальний аналіз. Прогнозний аналіз і часові ряди

Machine Learning and Data Mining. Predictive Analytics and Time Series Analysis

Лектор Чертов Олег Романович, доктор технічних наук, професор.

Тривалість курсу 30 аудиторних годин, 15 занять.

Програма дисципліни перебуває в процесі оновлення.

ПРОГРАМА КУРСУ

Заняття 1.

Вступ (мотивація).

- Визначення машинного навчання. Приклади задач різних типів навчання. Базова схема машинного навчання.

Заняття 2–3.

- Класифікаційні метрики: матриця помилок (Confusion matrix) і помилки першого та другого роду, доля правильних відповідей (Accuracy), точність (Precision) і повнота (Recall), F-міри (F-Scores), коефіцієнт кореляції Метьюса (Matthews Correlation Coefficient, MCC), Precision-Recall крива (curve), ROC (receiver operating characteristic)-крива, Area under curve (AUC), коефіцієнт Джині (Gini coefficient). Бінарний та багатокласовий випадки.

Заняття 4–5.

- Дерева прийняття рішень (Decision Trees). Ентропія Шеннона. Класифікація з дискретними та неперервними атрибутами. Виділення правил.

Заняття 6.

- Бінарні (CART) та небінарні (ID3, CHAID) види дерев. Критерії розщеплення: приріст інформації (Information Gain), нормалізований приріст інформації (Information Gain ratio), неоднорідність Джині (Gini impurity), помилка класифікації (Classification error), критерій Хі-квадрат (Chi-square test statistic for independence). Обрізання дерев (pre-pruning, post-pruning). Обробка пропущених значень, паралелізація.

Заняття 7.

- Метричні методи класифікації та регресії (Distance-Based Classification & Regression Methods). Жадібні та лінійні алгоритми (Eager vs Lazy learners). Гіпотези компактності та неперервності. Алгоритм к найближчих сусідів. Підбір вагів. Діаграма Вороного. Шумові характеристики. Нормалізація характеристик. Метрика Мінковського, косинусна міра, відстань Жаккарда, відстань Хемінга. Точні та наближені методи пошуку найближчих сусідів.

Заняття 8.

- Реалізація та дослідження методів класифікації та регресії (дерева прийняття рішень та метричні методи). Виконання індивідуальних домашніх завдань з реалізації та дослідження методів класифікації та регресії, презентація, обговорення і захист результатів.

Заняття 9.

- Лінійні моделі класифікації та регресії. Метод мінімізації емпіричного ризику. Поняття відступу (margin). Регресія зі штрафом. Різні види регуляризації. Сценарії Р-більше-N. Логістична регресія. Незбалансовані класи. Регресійні метрики.

Заняття 10.

- Баєсівські методи (Bayes methods). Безумовна (априорна) та умовна (апостеріорна) ймовірність. Теорема Баєса. Наївний Баєсів алгоритм (Naive Bayes Algorithm). Згладжування Лапласа.

Заняття 11.

- Реалізація та дослідження методів класифікації та регресії (логістична регресія та Баєсівські методи). Виконання індивідуальних домашніх завдань з реалізації та дослідження методів класифікації та регресії, презентація, обговорення і захист результатів.

Заняття 12.

- Кластеризація (Clustering). Типи кластерних алгоритмів: плоскі алгоритми (метод K-means) та алгоритми ієрархічної кластеризації (агломеративні та розділяючі). Жорстка кластеризація та м'яка (нечітка) кластеризація (метод C-means). Визначення кількості кластерів. Близькість кластерів. Моделі, засновані на щільності (DBSCAN і OPTICS).

Заняття 13.

- Реалізація та дослідження методів кластеризації. Виконання індивідуальних домашніх завдань з реалізації та дослідження методів кластеризації, презентація, обговорення і захист результатів.

Заняття 14.

- Відбір та синтез інформативних характеристик або зниження розмірності (Feature selection & extraction): відбір інформативних характеристик (feature selection), синтез інформативних характеристик (feature extraction), методи, вбудовані в модель навчання (embedded models). (Нормована) взаємна інформація. Метод головних компонент (principal component analysis, PCA). Метод незалежних компонент (independent component analysis, ICA). Лінійний дискримінантний аналіз (linear discriminant analysis, LDA).

Заняття 15.

- Реалізація та дослідження методів зниження розмірності. Виконання індивідуальних домашніх завдань з реалізації та дослідження методів відбору та/чи конструювання характеристик, презентація, обговорення і захист результатів.

Заняття 16.

- Комбінації моделей (Ensemble methods). Бутстреп-агрегування (Bootstrap method) або бергінг (bagging). Випадковий ліс (Random Forest). Boosting (підсилення). Алгоритм AdaBoost (адаптивне підсилення).

Заняття 17.

- Навчання з підкріпленням (Reinforced Learning). Базова модель навчання з підкріпленням. Марковський процес. Рівняння Беллмана. Алгоритм Q-навчання (Q-Learning).

Заняття 18.

- Реалізація та дослідження методів навчання з підкріпленням. Виконання індивідуальних домашніх завдань з реалізації та дослідження методів навчання з підкріпленням, презентація, обговорення і захист результатів.

Заняття 19.

- Асоціативні правила (Association rules). Виявлення елементів, що часто зустрічаються. Властивість антимонотонності підтримки. Замкнені набори елементів (Closed item sets). Рівень довіри (confidence). Алгоритм A-Priori. Алгоритм популярних наборів (Frequent-Pattern, FP-Growth). Підвищення ефективності обробки популярних наборів: метод локальних предметних наборів, семплінг.

Заняття 20.

- Причинно-наслідкове навчання (Causal Learning). Кореляція $\neq$ Причинність. Виявлення причинно-наслідкового зв'язку. Показник сили причинно-наслідкового зв'язку (Predictive Power Score, PPS).

Заняття 21.

- Навчання з гарантованою точністю (Conformal Learning). Класифікація та регресія із гарантованим навчанням (Conformal classification & regression). Функції невідповідності (Non-conformity functions). Застосування для випадку незбалансованих класів.

Заняття 22.

- Екзамен (Midterms).

Python для машинного навчання

Python for Machine Learning

Лектор	Подскребко Олександр Сергійович, кандидат економічних наук, доцент.
Тривалість курсу	18 аудиторних годин. 6 занять — по 3 години кожне заняття.

ПРОГРАМА КУРСУ

Заняття 1 (лекція/практичне заняття).

Основи роботи з бібліотекою Pandas.

- Pandas Series
- Pandas DataFrame
- Завантаження та робота з даними у Pandas

Заняття 2 (лекція/практичне заняття).

Візуалізація даних з Python.

- Побудова візуалізацій з matplotlib
- Побудова візуалізацій з Seaborn
- Бібліотека для візуалізації даних Plotly

Заняття 3 (лекція/практичне заняття).

Бібліотека Scikit-Learn. Основи Feature Engineering

- Підготовка даних до моделювання
- Обробка пропусків
- Feature Engineering

Заняття 4 (лекція/практичне заняття).

Pipeline для тренування моделей машинного навчання.

- Бібліотека Scikit-Learn для тренування моделей машинного навчання
- Model validation and evaluation

Заняття 5 (лекція/практичне заняття).

Підбір гіперпараметрів та аналіз отриманих результатів.

- Особливості процедури підбору гіперпараметрів
- GridSearchCV для підбору гіперпараметрів
- фреймворк Optuna для автоматизованого пошуку оптимальних гіперпараметрів для моделей машинного навчання
- Ensemble learning method

Заняття 6 (лекція/практичне заняття).

Ансамблі та аналіз результатів.

- Ensemble learning method - Random Forest
- Gradient boosting methods
- Аналіз результатів та підбір трешхолду

Нейронні мережі та Глибоке навчання

Neural networks and Deep Learning

Лектор Сергій Одрібець, к. ф.-м. н., Senior Data Scientist, Vodafone Ukraine

Тривалість курсу 24 аудиторні години
6 лекційних занять
9 практичних занять
4 лекційно-практичних заняття
1 контрольна робота/код рев'ю

ПРОГРАМА КУРСУ

Заняття 1 (лекція)

Вступ та розуміння Deep learning та Neural networks.

Заняття 2 (практичне заняття)

Neural networks на практиці, огляд популярних фреймворків, PyTorch та поняття tensor. Операції з tensor

Заняття 3 (практичне заняття)

Pipeline навчання Neural networks. Dataset та Dataloader

Заняття 4 (лекція)

Вступ в NLP (Natural Language Processing). Векторне представлення слів. Модель word2vec

Заняття 5 (практичне заняття)

Розуміння рекурентної нейронної мережі. Побудова архітектури RNN (Recurrent Neural Networks), LSTM та GRU на практиці

Заняття 6 (практичне заняття)

Згорткові нейронні мережі. Задачі комп'ютерного зору. Архітектура CNN (Convolutional Neural Networks) на практиці

Заняття 7 (контрольна робота/код рев'ю)

Code review

Інфраструктура та технології Big Data

Big Data Infrastructure and Technologies

Лектор Юрій Демченко (Yuri Demchenko).

Тривалість курсу 15 аудиторних годин, 5 занять.

Програма дисципліни перебуває в процесі актуалізації. Зміст і структура курсу можуть змінюватися.

ПРОГРАМА КУРСУ

Заняття 1. Вступ до курсу. Основи хмарних обчислень (Cloud Computing).

- Моделі хмарних сервісів, хмарні ресурси, принципи роботи сервісів, багатокористувацька архітектура (multitenancy).
- Віртуальні гібридні та динамічні хмарні центри обробки даних.
- Перенесення корпоративної IT-інфраструктури у хмару.
- Приклади використання хмарних технологій у бізнесі.
- Економіка хмарних сервісів і моделі ціноутворення.
- Практика: знайомство з Amazon Web Services (AWS), сервісами EC2 і S3, розгортання віртуальної машини та організація доступу.

Заняття 2. Хмарні технології та Big Data. Інфраструктура й алгоритми обробки великих даних.

- Визначення та властивості Big Data.
- Референсна архітектура Big Data.
- Приклади використання великих даних у науці та промисловості.
- Алгоритми обробки великих даних: MapReduce, Pregel та інші.
- Компоненти екосистеми Hadoop.
- Огляд основних хмарних платформ: AWS, Microsoft Azure, Google Cloud Platform (GCP).
- Практика: розгортання простого вебсайту з підключеною базою даних у середовищі AWS.
- Індивідуальне завдання: звіт про практичну роботу з AWS.
- Груповий проєкт: формування команд по 4–5 студентів і вибір теми проєкту на основі моделі підприємства.

Заняття 3. Платформа Hadoop, потокові дані та Apache Spark.

- Вступ до Hadoop і MapReduce.
- Екосистема Hadoop та її компоненти.
- Hadoop Distributed File System (HDFS).
- YARN, HBase, Hive та Pig.
- Інструменти програмування для Hadoop і MapReduce.
- Потоки даних і потокова аналітика.
- Інструменти потокової обробки: Kafka, Flume.
- Архітектура та компоненти Apache Spark.

- Популярні платформи Spark, зокрема Databricks.
- Програмування й інструменти Apache Spark.
- Практика: початок роботи з Databricks Community Edition, виконання демонстраційних вправ.
- Груповий проєкт: визначення даних для проєкту та вибір набору даних (dataset).

Заняття 4. Бази даних SQL та NoSQL.

- Повторення основ SQL.
- Типи баз даних NoSQL та їхні особливості.
- Колонкові бази даних і приклади їх використання, зокрема HBase.
- Сучасні масштабовані бази даних: AWS Aurora, Azure Cosmos DB, Google Spanner.
- Практика: експерименти з хмарними NoSQL-базами даних, зокрема MongoDB.
- Індивідуальне завдання: реалізація простої задачі MapReduce (наприклад, підрахунок слів) та алгоритму ранжування.
- Груповий проєкт: визначення інфраструктури Big Data для підприємства та потоків обробки даних.

Заняття 5. Управління великими даними та Data Governance.

- Архітектура управління даними (Data Management Architecture).
- Технічні компоненти управління даними.
- Корпоративні практики управління даними.
- DAMA DMBOK — Data Management Body of Knowledge.
- Практика: підготовка плану управління даними (Data Management Plan, DMP) для групового проєкту.
- Груповий проєкт: визначення структур даних, аналітичних застосунків і сервісів, вибір інфраструктурних компонентів.

Заняття 6. Отримання даних із вебджерел: Web API, Data Dumps та Web Scraping.

- Огляд доступних у мережі наборів даних.
- Методи вебскрейпінгу (web scraping).
- Принципи проєктування Web API.
- Специфікація OpenAPI.
- Практика: використання бібліотеки BeautifulSoup для збирання даних із вебсторінок.
- Груповий проєкт: реалізація простого алгоритму.

Заняття 7. Проміжний захист групових проєктів.

- Презентації проміжних результатів групових проєктів.
- Виступ кожної команди тривалістю приблизно 20 хвилин.
- Спільне обговорення результатів.
- Підготовка проміжного звіту відповідно до вимог проєкту.

Заняття 8. Amazon SageMaker як платформа для Data Analytics і Machine Learning.

- Функціональні можливості Amazon SageMaker.
- Сервіси платформи та приклади їх використання.
- Застосування SageMaker для аналітики даних і розробки моделей машинного навчання.
- Практика: підготовка плану управління даними (DMP) для проєкту.

- Груповий проєкт: продовження реалізації простого алгоритму.

Заняття 8. Управління проєктами Data Science. DataOps та MLOps.

- Моделі управління проєктами Data Science: CRISP-DM, ASUM, TDSP.
- Основи DataOps та MLOps.
- Практика: знайомство з Apache Pig, створення та виконання простих запитів.
- Індивідуальне завдання: використання мови Pig для програмування процесів обробки великих даних.
- Груповий проєкт: вибір інфраструктурних сервісів і компонентів.

Заняття 9. Безпека великих даних і відповідність нормативним вимогам.

- Основні виклики безпеки Big Data.
- Захист даних.
- Управління доступом та ідентифікацією користувачів (Identity Management).
- Переваги федеративної інфраструктури контролю доступу.
- Хмарні сервіси забезпечення безпеки.
- Стандарти відповідності (compliance).
- Оцінювання сервісів хмарних провайдерів щодо відповідності вимогам безпеки.
- Практика: перевірка відповідності хмарної інфраструктури вимогам безпеки.
- Груповий проєкт: питання безпеки, нормативної відповідності та захисту даних.

Заняття 10. Фінальний захист групових проєктів.

- Презентації групових проєктів.
- Обговорення результатів та отриманого досвіду.
- Завершення й подання групових проєктів з інфраструктури Big Data.

Практичний модуль курсу з інфраструктури: «Від локального data product до хмари»

From Local Data Product to the Cloud

Лектор	Рижук Володимир Володимирович, Senior Data Analyst / Data Scientist, Nova Poshta.
Тривалість курсу	9 аудиторних годин. 3 практичних заняття по 3 години кожне.

ПРОГРАМА КУРСУ

Заняття 1. Продукт у хмарі: Supabase + Vercel

- Демонстрація: dev і prod середовища, той самий код з різними рядками підключення (DATABASE_URL).
- Перевірка стану групових проєктів після курсу DWH (запуск docker compose з нуля).
- Практика: створення prod-бази в Supabase, накочування міграцій, запуск ETL проти prod.
- Практика: публікація PWA на Vercel, окрема роль бази тільки для читання, встановлення PWA на телефон.
- Результат: продукт з реальними даними доступний з будь-якого пристрою.

Заняття 2. Автоматизація: CI/CD і нічний пайплайн

- Демонстрація: GitHub Actions як оркестратор, тести даних як тести коду.
- Практика: workflow на кожну зміну коду — збірка Docker-образу, тимчасова база, міграції, тести якості даних (унікальність, NOT NULL, кількість рядків).
- Практика: нічний ETL за розкладом проти prod-бази, секрети в GitHub Secrets, оновлення вітрин і вивантаження даних для Tableau.
- Практика: окремі preview-розгортання фронтенду для кожної зміни.
- Результат: дані оновлюються щодня без участі людини, а помилковий код не потрапляє в prod.

Заняття 3. Готовність до експлуатації, альтернатива в Google Cloud та захист

- Демонстрація: той самий продукт у Google Cloud (Cloud Run, Cloud Scheduler, PostgreSQL), основи інфраструктури як коду (Terraform), порівняння варіантів за вартістю та складністю.
- Практика: алерт про збій пайплайну, відображення свіжості даних у PWA, резервна копія бази та перевірка відновлення.
- Практика: архітектурний документ (ADR) — вибір варіанта, оцінка вартості, план масштабування «×100».
- Захист групових рішень (5–7 хвилин на групу), підсумки модуля.

Сучасні технології штучного інтелекту та їх практичне застосування

AI for Data Professionals & Applied AI Systems

Лектор	Кукса Сергій, Senior Data Scientist, AI practitioner, MSc in Technology Leadership & AI.
Тривалість курсу	24 аудиторні години. 8 занять по 3 години.

ПРОГРАМА КУРСУ

Розділ I. AI for Data Professionals

Тривалість: 12 аудиторних годин, 4 заняття по 3 години.

Модуль 1. AI Foundations & Practical AI Tools for Data Professionals

- Огляд еволюції штучного інтелекту та сучасного AI landscape. Співвідношення понять Artificial Intelligence, Machine Learning, Deep Learning, Generative AI, Large Language Models, multimodal AI та AI Agents.
- Аналіз впливу Generative AI та LLM на сучасні data-oriented професії. Основні сценарії використання AI у роботі Data Analyst, Data Scientist та Data Engineer.
- Можливості та обмеження сучасних AI-моделей. Hallucinations, reliability, privacy, confidentiality та залежність якості відповіді від контексту.
- Основи Prompt Engineering. Формування структурованого prompt із використанням context, task, constraints та output format. Огляд основних параметрів генерації та підходів до взаємодії з LLM.
- Огляд сучасних AI-інструментів: chat applications, AI assistants у робочих середовищах, AI для документів, таблиць і презентацій, API, cloud та local AI solutions.
- Google Colab як основне середовище практичної роботи. Базові приклади генерації, пояснення та виправлення Python-коду за допомогою AI.

Модуль 2. AI as a Learning & Problem-Solving Assistant

- Основи програмної взаємодії з Large Language Models. Поняття API, API key, request, response, model, tokens та context.
- Виклик LLM через Python у Google Colab. Побудова базових функцій для багаторазової взаємодії з AI-моделями та огляд можливості використання різних AI providers.
- Передача моделі власного контексту та використання AI для роботи з навчальними матеріалами, текстами, транскриптами та відео.
- Автоматичне формування summary, structured notes, glossary, key concepts, quiz questions та інших навчальних матеріалів.
- Побудова простого AI Learning Assistant для відповідей на питання на основі заданого контенту.
- Практичні методи використання AI для пояснення складних концепцій, self-testing, виявлення прогалин у знаннях та адаптації пояснення до рівня користувача.
- Принципи отримання перевірюваних пояснень, припущень, формул і проміжних результатів під час розв'язання логічних та математичних задач.

Модуль 3. AI for Python & Data Work

- Використання AI як coding assistant під час роботи з Python.
- Генерація та пояснення Python-коду, debugging, code review, refactoring та пошук помилок за допомогою AI.
- AI-assisted робота з Pandas: завантаження, дослідження, очищення та трансформація даних.
- Аналіз missing values, duplicates, incorrect data types, аномалій та інших типових проблем якості даних.
- Використання AI для exploratory data analysis та формування можливих напрямів подальшого аналізу.
- Інтеграція LLM через API у Python data workflow. Передача schema, sample data, descriptive statistics та результатів розрахунків для подальшої інтерпретації.
- Робота зі structured outputs та використання результатів LLM у програмному процесі.
- Особлива увага приділяється розмежуванню синтаксично коректного коду та коректної бізнес-логіки, а також методам перевірки AI-generated code.

Модуль 4. AI + SQL + Analytics

- Використання AI для роботи з SQL та аналітичними задачами.
- Перехід від natural-language business question до SQL-запиту. Принципи Text-to-SQL.
- Використання AI для schema exploration, аналізу структури таблиць та визначення зв'язків між ними.
- Генерація та пояснення SQL-запитів із використанням SELECT, WHERE, GROUP BY, JOIN, CTE та window functions.
- AI-assisted debugging, покращення структури та читабельності SQL-коду.
- Використання LLM API для програмної генерації SQL та виконання запиту в аналітичному середовищі.
- Побудова базового AI-assisted analytical workflow:
Business Question → AI → SQL → Database → Result → AI → Business Explanation
- Використання AI для exploratory analytics, формування structured analytical outputs, документації та synthetic test data.
- Перевірка відповідності технічно правильного SQL реальній бізнес-логіці поставленої задачі.

Розділ II. Applied AI Systems

Тривалість: 12 аудиторних годин, 4 заняття по 3 години.

Модуль 5. From Neural Networks to Foundation Models

- Перехід від класичних Neural Networks до сучасних Foundation Models. Зв'язок між уже вивченими принципами neural networks та архітектурою сучасних Large Language Models.
- Представлення текстової інформації: tokenization, tokens, vocabulary та embeddings. Поняття vector representations та semantic similarity.
- Короткий огляд sequence modelling та еволюції від RNN/LSTM до механізму Attention.
- Принципи роботи Self-Attention та Transformer architecture. Роль embeddings, positional information, attention та feed-forward layers.
- Pre-training, next-token prediction, instruction tuning та alignment. Відмінність між base model, instruction-tuned model та сучасними multimodal Foundation Models.
- Принципи inference: context window, token generation, sampling, temperature та non-deterministic behaviour. Причини hallucinations та обмеження LLM.
- Знайомство з LangGraph як середовищем для побудови AI workflows: поняття State, Node та Edge.

- Початок наскрізного проєкту ML Model Companion: використання LLM для перетворення ML prediction та структурованих customer data у зрозуміле business explanation.
- Робота зі structured outputs та розмежування понять customer facts, ML prediction та explanation. Принцип: LLM може пояснювати надані evidence, але не повинна самостійно вигадувати причини prediction.

Модуль 6. LLM Applications, RAG & AI Assistants

- Перехід від Foundation Model до повноцінного AI application. Відмінність між model та application layer, що включає data, retrieval, business rules та orchestration.
- Обмеження підходу передачі всієї інформації безпосередньо в prompt: context window, token cost, latency, relevance та підтримка актуальності knowledge base.
- Практичне застосування embeddings для semantic search. Поняття query embedding, document embedding, similarity search та top-k retrieval.
- Принципи підготовки knowledge base: chunking, chunk size, overlap, semantic boundaries та metadata.
- Архітектура Retrieval-Augmented Generation (RAG): document ingestion, embeddings, vector store, retrieval, context augmentation та generation.
- Відмінність між RAG та Fine-Tuning. Використання RAG для додавання актуальної або domain-specific knowledge без зміни weights моделі.
- Grounding, source-based answering, citations та робота з випадками, коли необхідної інформації немає в knowledge base.
- Розширення LangGraph workflow: додавання retrieval layer та побудова RAG-enabled Model Assistant на основі model card, feature dictionary, model metrics, limitations та business glossary.
- Інтеграція ML explainability із RAG та LLM. Формування grounded business explanation на основі реальних model outputs, feature contributions і документації.
- Практичний принцип: ML predicts → Explainability provides evidence → RAG provides knowledge → LLM communicates.

Модуль 7. AI Agents, Evaluation & Production Systems

- Перехід від AI Assistant до Agentic System. Відмінність між LLM, Tool Calling, Workflow та AI Agent.
- Поняття tools та function calling. Використання Python functions, APIs, databases, ML models та retrieval systems як зовнішніх інструментів для LLM.
- Побудова agentic workflow у LangGraph. State management, tool nodes, conditional routing та termination conditions.
- Порівняння deterministic workflows та agentic approaches. Визначення випадків, коли Agent додає цінність, а коли простіший workflow є більш надійним рішенням.
- Основні компоненти AI Agent: model, instructions, tools, state, memory, routing, guardrails та human-in-the-loop.
- Короткий огляд single-agent та multi-agent architectures. Роль planner/executor та supervisor/worker patterns.
- Розширення ML Model Companion tools для отримання customer profile, ML scoring, prediction explanation та пошуку model documentation.
- Побудова Agent, який самостійно визначає, які tools необхідно використати для відповіді на business request.
- Основи evaluation AI systems: tool selection accuracy, argument correctness, numeric accuracy, groundedness, relevance та task completion.
- Порівняння deterministic evaluation, human evaluation та LLM-as-a-Judge.

- Production considerations: reliability, retries, latency, token usage, cost, logging, observability, privacy, access control, prompt injection, guardrails та human approval.

Модуль 8. AI for Capstone: From ML Model to AI-Enabled Solution

- Перехід від технічної ML-моделі до завершеного business solution. Розмежування понять model, prediction, interpretation, decision, action та business outcome.
- Принцип Start with the Business Problem: визначення користувача рішення, decision process, data requirements та ролі ML/AI у бізнес-процесі.
- Вибір архітектури залежно від задачі: Classical ML, LLM, RAG, Workflow, Tools, Agent або hybrid solution.
- Визначення ситуацій, коли Generative AI не потрібен. Принцип використання найпростішої архітектури, яка достатньо вирішує business problem.
- Model explainability та AI-generated explanation. Розмежування technical evidence і natural-language communication.
- Перехід від Model Metrics до Business Metrics. Інтерпретація Accuracy, Precision, Recall, F1 та ROC AUC з погляду operational та business impact.
- Human-in-the-loop та визначення відповідного рівня automation залежно від ризику, вартості помилки та impact на користувача.
- Практичний варіант code-first implementation: Streamlit + LangGraph. Побудова простого business interface для ML Model Companion із можливістю scoring, explanation та Q&A.
- Практичний варіант low-code implementation: n8n workflow. Побудова AI workflow з доступом до ML scoring service, customer data та model documentation.
- Порівняння code-first та low-code approaches і вибір підходу залежно від рівня customisation, complexity та задачі.
- Використання Generative AI для підготовки Capstone presentation, формування business-friendly explanations та проведення AI Mock Defence перед фінальним захистом.
- Опціональне розширення Final Project: від AI-assisted presentation до RAG Model Assistant або повноцінного Agent, який може використовувати власну ML-модель студента.

Лектор	Петруня Руслан
Тривалість курсу	9 аудиторних годин. 3 практичні воркшопи по 3 години.

ПРОГРАМА КУРСУ

Воркшоп 1. Хмари та віртуалізація

Хмарні технології давно стали звичною частиною ІТ-інфраструктури, але їх часто плутають із віртуалізацією або віддаленим зберіганням даних. Воркшоп присвячений принципам Cloud та відмінностям хмарної моделі від класичної інфраструктури.

- ITIL 4 як підхід до управління ІТ-сервісами.
- CISC та RISC у контексті архітектури обчислювальних систем.
- Роль абстракції в інфраструктурних рішеннях.
- Відмінності між IaaS, PaaS та SaaS.
- П'ять базових ознак хмарної моделі.
- Межа між віртуалізацією та хмарною інфраструктурою.

Воркшоп 2. Основи Git та GitHub для Data Analyst

Практичний воркшоп із роботи з Git та GitHub для аналітиків даних, науковців з даних та всіх, хто працює з Python/SQL-скриптами.

- Базові принципи контролю версій.
- Створення та ведення локального репозиторію.
- Коміти та історія змін.
- Гілки та їх використання.
- Публікація проєкту на GitHub.
- Командна робота через pull requests.
- Python-скрипти, SQL-файли, ноутбуки та структура аналітичного проєкту.

Воркшоп 3. Контейнери

Тривалість: 3 години.

Програма воркшопу перебуває в процесі розробки. Детальний опис буде додано після затвердження.