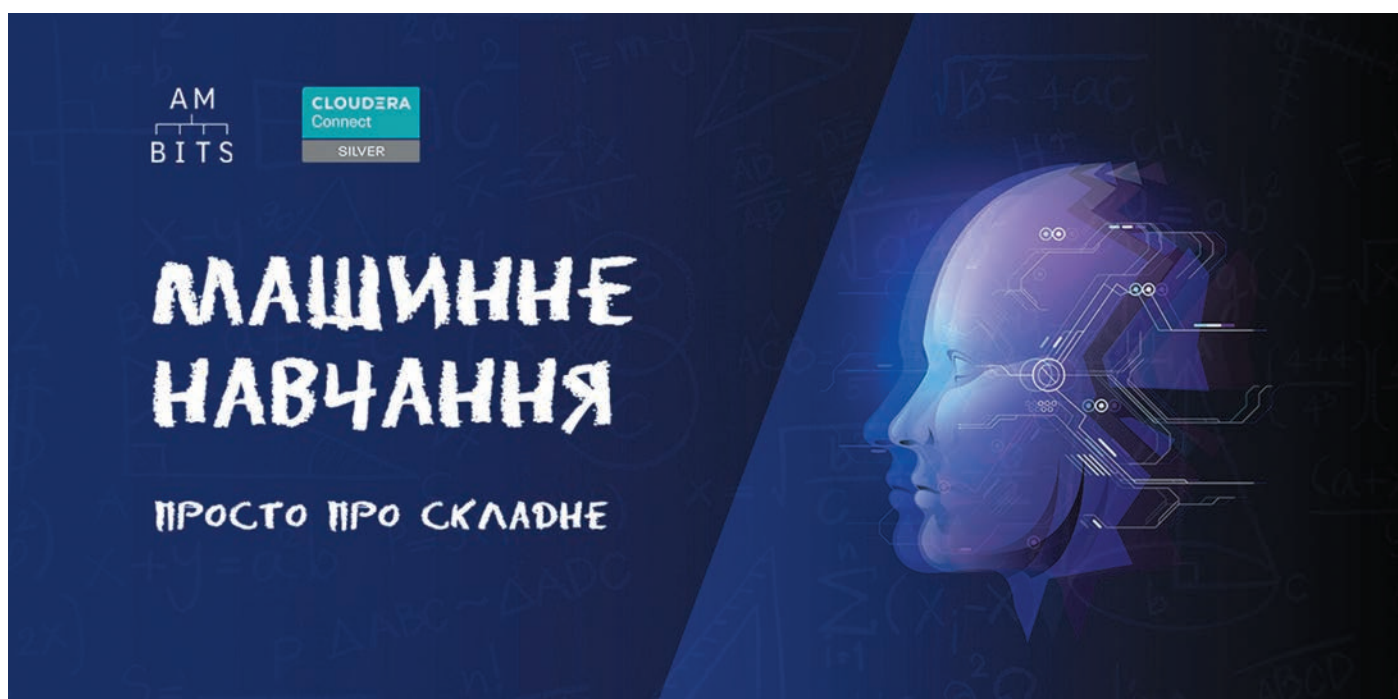


ML-технології навколо нас



Термін Machine Learning (машинне навчання) посів важливе місце в сьогоднішній — як у трендах новин, так і на ринку праці у сфері автоматизації. Проте ML залишається досить складною темою через свою всеосяжність, новизну та високі темпи розвитку, і залишає безліч питань. Спробуємо розібратись, що таке машинне навчання, дамо основні визначення та поділимося досвідом щодо прикладних завдань, які можна вирішувати за допомогою ML.

Машинне навчання — це набір методів у галузі штучного інтелекту, що їх застосовують для створення моделі, яка навчається на певному наборі даних. В процесі навчання модель обробляє різноманітні масиви вхідних даних і знаходить у них закономірності. Для побудови таких моделей використовуються засоби математичної статистики, чисельні методи, математичний аналіз, методи оптимізації, теорія ймовірності, нейронні мережі та інші техніки роботи з даними у цифровій формі.

Якщо узагальнити, ML — це інструмент, за допомогою якого вирішується певний клас задач, пов'язаних з необхідністю виявити закономірності у складних багатопараметричних завданнях, що не можуть бути вирішені класичними методами через надто велику кількість параметрів чи неочевидність їх пов'язаності.

3 чого все починається

Розробка ML-рішення — це комплексний процес, що вимагає взаємодії різних систем та навичок багатьох фахівців. Наприклад, для збирання та збереження даних відповідний експерт — Data Engineer — розробляє ETL-процеси та взаємодіє з базами даних. Інший фахівець — Data

Analyst — проводить аналіз даних, шукає закономірності та взаємозв'язки, перевіряє статистичні гіпотези. Разом з ними працює ML-інженер, який розробляє модель рішення, експериментує з різними його архітектурами, шукає оптимальні параметри для отримання найкращого результату.

З 2016 року компанія **AM-BITS** успішно реалізує проекти на основі технологій Big Data, AI та ML, а поштовхом до цього стало партнерство з американським розробником **Hortonworks**, що є одним з лідерів галузі. З 2019 року AM-BITS отримав статус «срібного» партнера компанії **Cloudera**. Наразі AM-BITS об'єднує досвідчену та сертифіковану команду, що складається з понад 25 експертів з роботи з корпоративними даними. Компанія має досвід

у галузі обробки та дослідження даних, зокрема за такими напрямками, як розробка корпоративних платформ даних, впровадження рішень в області обробки та аналізу даних, побудова моделей машинного навчання для технічних і бізнес-задач в різних секторах економіки (фінанси, телеком, медіа).

Під час роботи над ML-проектами експерти AM-BITS надають перевагу платформі Cloudera Data Science Workbench. Завдяки інтегрованій в CDP кластер CDSW платформі різні фахівці мають можливість ефективно реалізовувати проекти в галузі ML і Data Science (аналіз, обробка та надання даних для аналітичних інструментів) (рис. 1).

Відповідно до моделі CDSW розробка кожної ML-моделі ведеться в окремому проекті, де

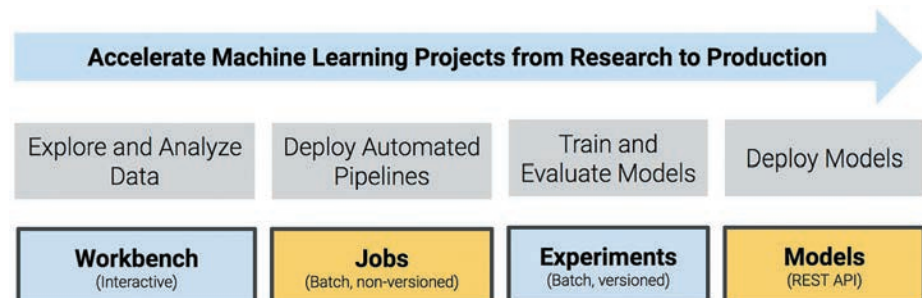


Рис. 1. Етапи створення ML-проекту

можуть взаємодіяти члени команди з різними правами доступу залежно від завдань. Таким чином, можна вести розробку проєктів на загальній платформі різними командам (розробникам), при цьому дотримуючись ізольованості один від одного, що дозволяє співробітникам працювати паралельно, використовуючи тільки необхідні ресурси.

Машинне навчання для українських замовників

Практичне застосування ML-проєктів досить широке. Так, наприклад, інженери AM-BITS розробили рішення для прогнозування генерації сонячної електроенергії, а також систему для розрахунку очікуваного рівня її споживання. Для побудови точних прогнозів використовувалися історичні дані щодо генерації/споживання електроенергії, які було збагачено інформацією щодо прогнозу погоди та інших відкритих даних. Завдяки використанню CDSW вдалося автоматизувати збір необхідних погодних даних, підбір архітектури моделі та гіперпараметрів. Метою обох проєктів було одержання короткострокового прогнозу для планування закупівель на українській енергетичній біржі.

Ось інший приклад. Фахівці компанії AM-BITS, зокрема, розробили систему ідентифікації особи за обличчям – Face ID (рис. 2). Інженери компанії скористалися останніми дослідженнями у сфері розпізнавання обличчя і на їх основі побудували власне рішення. Модифікувавши та оптимізувавши певні існуючі алгоритми в рамках вирішення поставленої задачі, фахівці отримали модель, яка дозволяє з високою точністю розпізнавати обличчя людей, навіть якщо частина лиця прихована, наприклад, респіраторною маскою.

Серед цікавих проєктів можна згадати розробку системи біометричної верифікації клієнтів за голосовими даними.

Задля ефективного порівняння голосових відбитків було побудовано нейромережу (DNN, Deep Neural Network) модель, що навчається за допомогою бібліотеки PyTorch. Цьому етапу

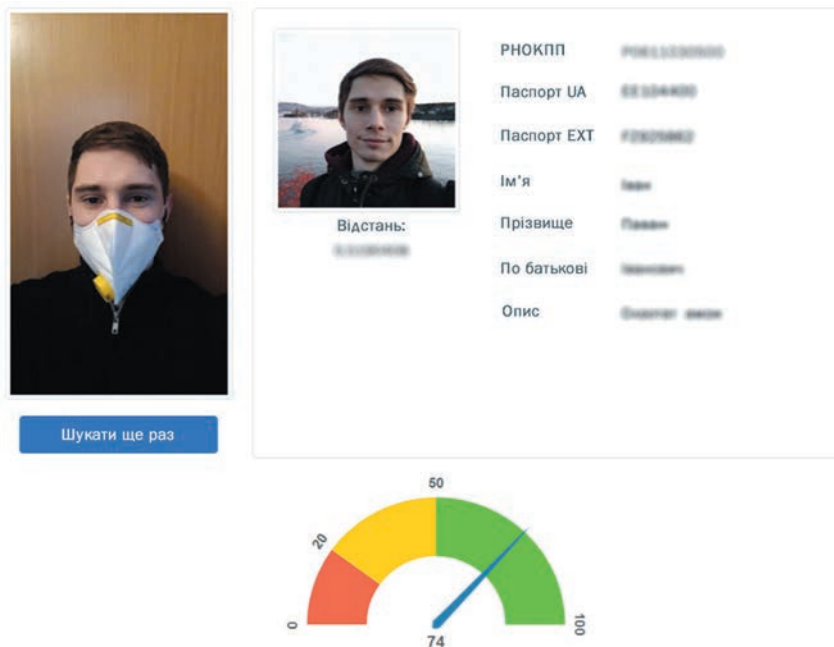


Рис. 2. Система ідентифікації особи за обличчям — Face ID

передували отримання голосових відбитків та їх ідентифікація за набором мел-кепстральних коефіцієнтів. Ідентифікація особи за голосом може бути використана для пришвидшення обслуговування в контакт-центрі, миттєво надаючи оператору дані щодо абонента, а також попередню історію комунікацій з метою покращення клієнтського досвіду.

Але можливості машинного навчання не обмежуються прогнозуванням чи ідентифікацією.

Наразі більшість інформації – новини, аналітика, прогнози тощо – створюється та споживається у відеоформаті. Ми створюємо безліч інформаційних повідомлень, фіксуємо щоденні події та ділимося ними з аудиторією, але знайти повторно корисну або просто цікаву інформацію часом надзвичайно складно або навіть неможливо.

З метою полегшення роботи численних аналітиків експерти AM-BITS розробили багатондову лінійно-масштабовану платформу для автоматичного аналізу відео за допомогою

апаратних графічних процесорів. Для цього були створені моделі машинного навчання з використанням згорткових нейронних мереж (Convolutional Neural Networks) і додаткових моделей оптимізації та відстеження об'єктів. Для швидкої підготовки датасету, необхідного для навчання моделей, було введено додаткову підсистему відеорозмітки. Крім того, розроблено інтерфейс для ведення відеотеки, її перегляду та аналізу, функціонал дозволяє завантажувати відео та обирати моделі аналізу для його обробки (рис. 3).

Автоматичне тегування та аналіз відеофайлів дозволяють не лише швидко знайти необхідну інформацію серед переглянутих матеріалів, але й проаналізувати та позначити необхідну інформацію в безперервному відеопотоці, що полегшує роботу експертів, аналітиків, журналістів та інших зацікавлених осіб. Ідентифікацію особи за фото Face ID можна використовувати для швидкого та безпечного підтвердження особи при використанні різних додатків або для запобігання шахрайству та підробленню документів.

Також цей інструмент стане в нагоді у справі виявлення колабораціоністів, бо тегування та аналіз відеоматеріалів можна використовувати не лише для обробки стрічки новин, але і для виявлення на відео військових злочинців з метою подальшого притягнення винних до відповідальності.

Підсумовуючи, варто зазначити, що ML перетворюються на звичний інструмент для вирішення поточних завдань, і команда AM-BITS допомагає компаніям та організаціям досягати поставлених цілей, зокрема з допомогою цих потужних технологій.

Підготовлено за матеріалами компанії **AM-BITS** <https://am-bits.com/>

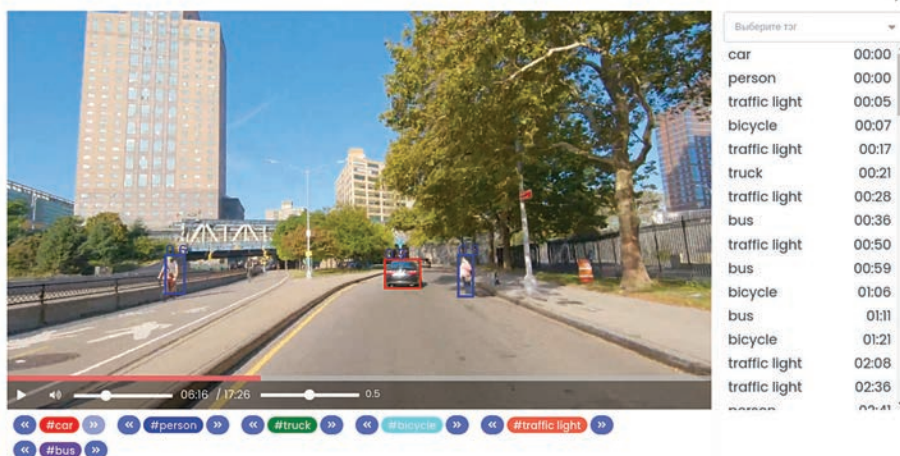


Рис. 3. Аналіз графічних об'єктів у відеопотоці