1. Introduction to Machine Learning & Data

Thursday, 8 January 2026 10:21 AM

Introduction to Machine Learning

What is Machine Learning?

Machine Learning (ML) is a field of computer science that uses **statistical techniques** to give computer systems the
ability to **learn from data** and improve their performance **without being explicitly programmed**.

Instead of writing rules manually, we **train the system using data**, and the system learns the logic on its own.

What is Explicit Programming?

Explicit Programming means:

- For **every possible scenario**, the developer writes **code manually**
- Logic is fixed using **if–else conditions**
- If a new case appears, the **code must be changed**

Example:
If condition A → do X
If condition B → do Y

This approach works only when:

- Scenarios are limited
- Rules do not change frequently

Where Can We Use Machine Learning?

When You Cannot Write Code for Every Scenario

Some real-world problems are:

- Too complex
- Continuously changing
- Impossible to define with fixed rules

Example 1: Email Spam Detection

Traditional (Rule-Based) Approach

A software developer may write rules like:

- If email contains **many images**
- If the word **“discount” appears more than 3 times**
- If words like **“awesome”, “offer”** are used
- If email has **very large text or links**
- Then mark email as **Spam**

Problem with Rule-Based Approach

- **Advertising companies learn these rules**

- They modify emails to avoid spam filters
- Example:
    - Change wording
    - Reduce images
    - Alter email structure

Result:

- Program **fails to detect spam**
- Developer must **change logic again**
- This cycle repeats again and again

How Machine Learning Helps Here

- ML **learns from past email data**
- It understands **patterns**, not fixed rules
- When new spam types appear:
    - We provide **new data**
    - Model **updates its logic automatically**

Key Point:

If data changes, **ML logic also changes automatically**, without rewriting code.

Example 2: Image Classification (Dog Identification)

Suppose we want to identify **dogs in images**.

Explicit Programming Problem

- How many dog images can we write rules for?
- Dogs look different:
    - Different colors
    - Different sizes
    - Different breeds
    - Different angles

Writing rules for every image is **impossible**.

Machine Learning Solution

- Train the model with **thousands of dog images**
- Model learns:
    - Shape
    - Texture
    - Patterns
- Then it can identify **new dog images** it has never seen before

Important Use Case: Data Analysis vs Data Mining

Data Analysis

- Used to:
    - Analyze existing data
    - Identify **known patterns**
    - Generate reports and outputs

Example:

- Monthly sales report
- Average marks of students

Data Mining

- Used to:
  - Discover **hidden patterns**
  - Find relationships we are **not explicitly looking for**
- These patterns **cannot be easily found** using normal data analysis

Example:

- Customer buying behavior
- Fraud detection patterns

History of Machine Learning

- Machine Learning **existed 40–50 years ago**
- It came early in the industry
- But it was **not popular earlier**

Why ML Was Not Popular Before?

1. **Data collection problem**
   - Data was not easily available
2. **Hardware limitation**
   - Low memory
   - Slow processors
   - Expensive storage

Why ML Became Popular After 2010?

- Internet evolution
- Smartphone revolution
- Social media and apps generating huge data
- Powerful hardware available
- Big data centre

Today:

- Even mobile phones have **up to 128 GB RAM**
- Earlier, even servers did not have this capacity

Conclusion

Machine Learning is used when rules are hard to define, data keeps changing, and systems need to learn
automatically from experience.