Machine Learning – Class 02

A Phitron AI/ML Series Presentation

by Md. Sabbir Ahmed

Today's Learning Journey: Our Agenda

Embark on an insightful exploration into the realm of Machine Learning, understanding its core principles and diverse applications.

01

Deciphering Machine Learning

Understanding what ML entails and how it differs from broader AI and specific Deep Learning concepts.

02

Real-World Implementations

Exploring prevalent applications of ML that shape our daily lives and industries.

03

Categorising ML Approaches

Delving into the primary types of Machine Learning: Supervised, Unsupervised, and Reinforcement Learning.

04

The ML Training Pipeline

A step-by-step breakdown of the complete process, from problem definition to model deployment.

05

Career Pathways in ML

Navigating the exciting opportunities and essential skills required for a career in this field.

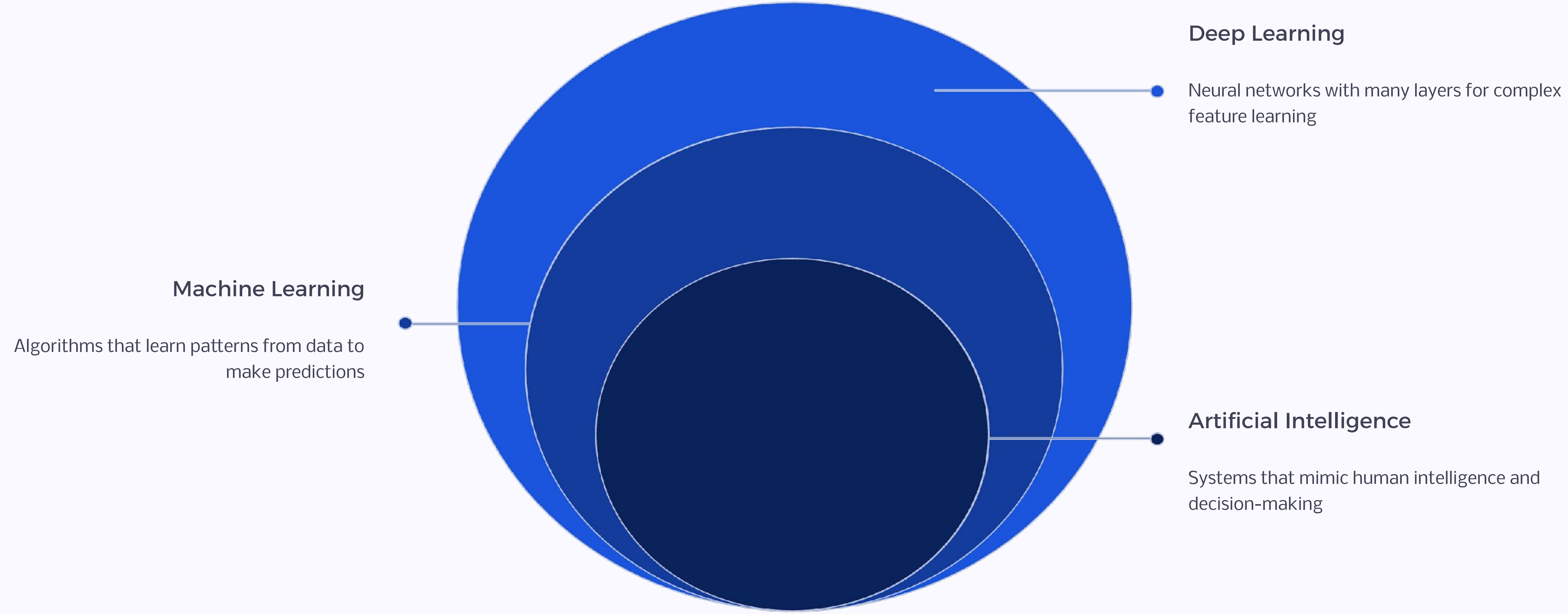
06

Concluding Remarks & Q&A

Recap of key takeaways, open discussion, and a sneak peek into our next session.

What Exactly is Machine Learning?

Machine Learning (ML) empowers machines to learn directly from data, identifying patterns and making decisions without explicit programming. It's a fundamental shift in how we approach problem-solving with computers. It's important to understand how ML fits within the broader field of Artificial Intelligence (AI), with Deep Learning (DL) being a specialized subset of ML.



What Exactly is Machine Learning? (Continued)

The Human Analogy

To better grasp these concepts, consider the human brain as a powerful analogy:

Artificial Intelligence (AI): The 'brain' itself, representing the overall cognitive capabilities to reason and solve problems.

Machine Learning (ML): The 'learning process' within the brain, acquiring knowledge from experiences and data.

Deep Learning (DL): The 'neurons' and their intricate connections, forming the foundational architecture that enables complex learning.

Machine Learning in Action: Real-World Applications

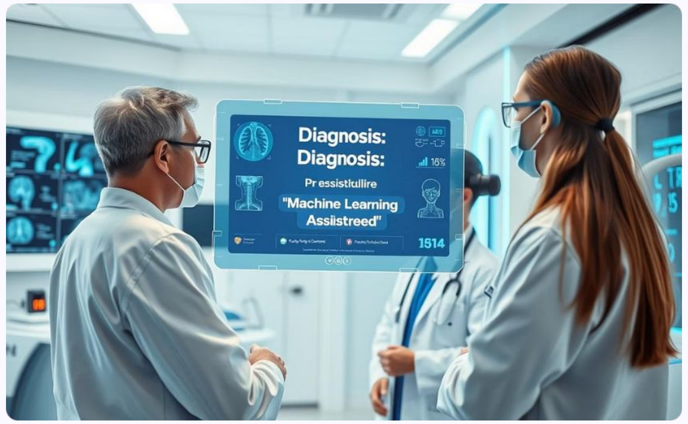
Machine Learning is not a futuristic concept; it's already deeply integrated into numerous aspects of our daily lives, quietly powering many of the services and technologies we rely upon.

Spam Email Filtering

Sophisticated ML algorithms analyse incoming emails to identify and quarantine unwanted spam, protecting your inbox from clutter and malicious content.

Content Recommendations

Streaming platforms like Netflix and YouTube use ML to predict your preferences, suggesting films, shows, and videos you're most likely to enjoy, enhancing your user experience.



Healthcare Diagnostics

ML models assist medical professionals in early disease detection, analysing images (like X-rays or MRIs) and patient data to provide accurate and timely diagnoses.

Fraud Detection in Finance

Banks employ ML to swiftly identify anomalous transaction patterns, flagging potential fraud in real-time and safeguarding your financial assets.

Machine Learning in Action: Real-World Applications (Continued)

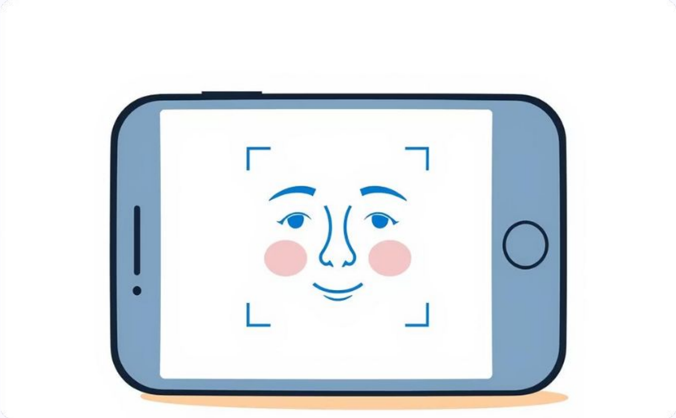
Machine Learning is not a futuristic concept; it's already deeply integrated into numerous aspects of our daily lives, quietly powering many of the services and technologies we rely upon.

Autonomous Vehicles

Self-driving cars leverage advanced ML and Deep Learning to perceive their environment, navigate, and make driving decisions, promising safer and more efficient transport.

Voice Assistants

Devices like Siri, Alexa, and Google Assistant use ML for natural language processing, allowing them to understand and respond to spoken commands and questions.



Facial Recognition

From unlocking your smartphone to enhanced security systems, ML-powered facial recognition identifies individuals by analysing unique facial features.

Personalized Marketing

E-commerce sites and advertisers use ML to analyse browsing history and purchase patterns, delivering highly relevant product recommendations and advertisements.

Categorising Machine Learning Approaches

Machine Learning is broadly categorised into three main types, each designed to solve different kinds of problems by learning in distinct ways.

Supervised Learning

This type involves learning from a labelled dataset, where each input has a corresponding correct output. The model learns to map inputs to outputs.

Classification: Predicting discrete categories (e.g., spam or not spam, cat or dog).

Regression: Predicting continuous values (e.g., house prices, temperature forecasts).

Unsupervised Learning

Here, the model works with unlabelled data, seeking to discover hidden patterns or structures within the dataset without explicit guidance.

Clustering: Grouping similar data points together (e.g., customer segmentation).

Dimensionality Reduction: Reducing the number of features while retaining important information (e.g., for visualisation or efficiency).

Reinforcement Learning

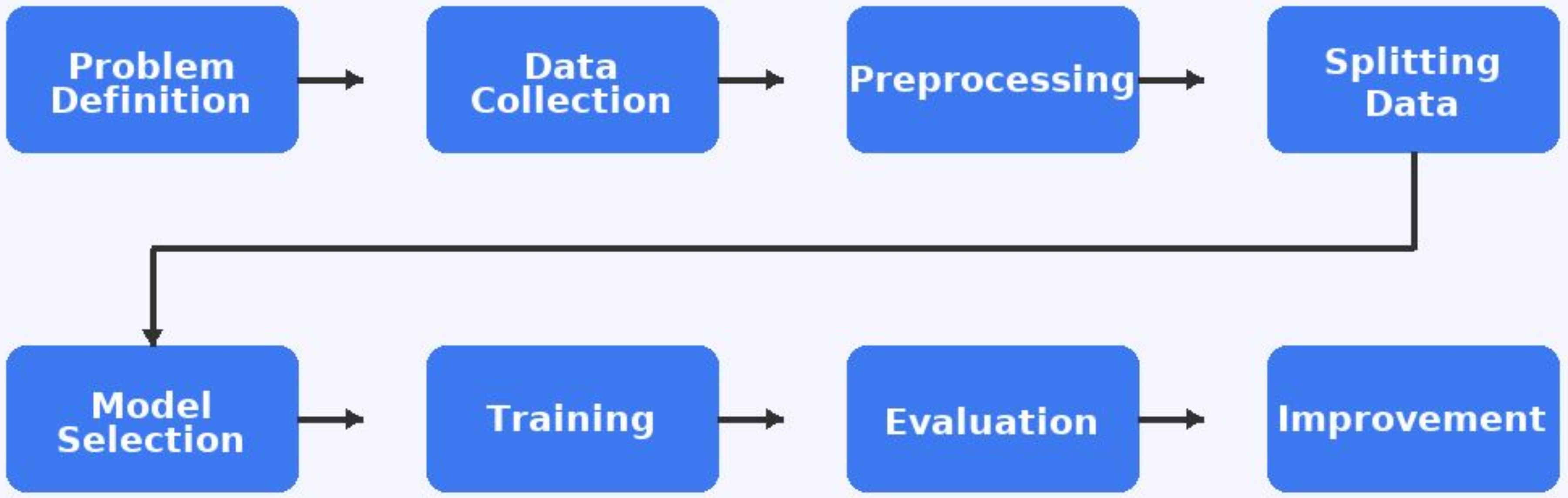
Inspired by behavioural psychology, an agent learns to make decisions by performing actions in an environment and receiving rewards or penalties.

AlphaGo: Google's programme that mastered the game of Go, defeating human champions.

Robotics: Training robots to perform complex tasks through trial and error in physical or simulated environments.

The ML Training Pipeline: A Systematic Approach

Developing a robust Machine Learning model involves a structured, multi-step pipeline, ensuring data quality, effective training, and reliable performance.



Problem Definition

Clearly articulate the problem to be solved and define measurable objectives.

Data Collection

Gather relevant and sufficient data from various sources. The quality and quantity of data are paramount.

Preprocessing

Clean, transform, and prepare data for model consumption. This includes handling missing values, normalisation, and feature engineering.

Splitting Data

Divide the dataset into training, validation, and test sets to ensure unbiased model evaluation.

Model Selection

Choose the appropriate ML algorithm based on the problem type and characteristics of the data.

Training

Feed the training data to the chosen model, allowing it to learn patterns and adjust its internal parameters.

Evaluation

Assess the model's performance using unseen test data and relevant metrics (e.g., accuracy, precision, recall).

Improvement

Iteratively refine the model through hyperparameter tuning, feature engineering, or trying different algorithms until desired performance is achieved.

Case Study: The Iris Dataset

A classic dataset in Machine Learning, the Iris dataset is often used to demonstrate classification algorithms. It comprises measurements of different parts of Iris flowers.

Dataset Features

The dataset contains 150 samples of Iris flowers, with four features measured in centimetres for each sample:

Sepal Length: The length of the sepal.

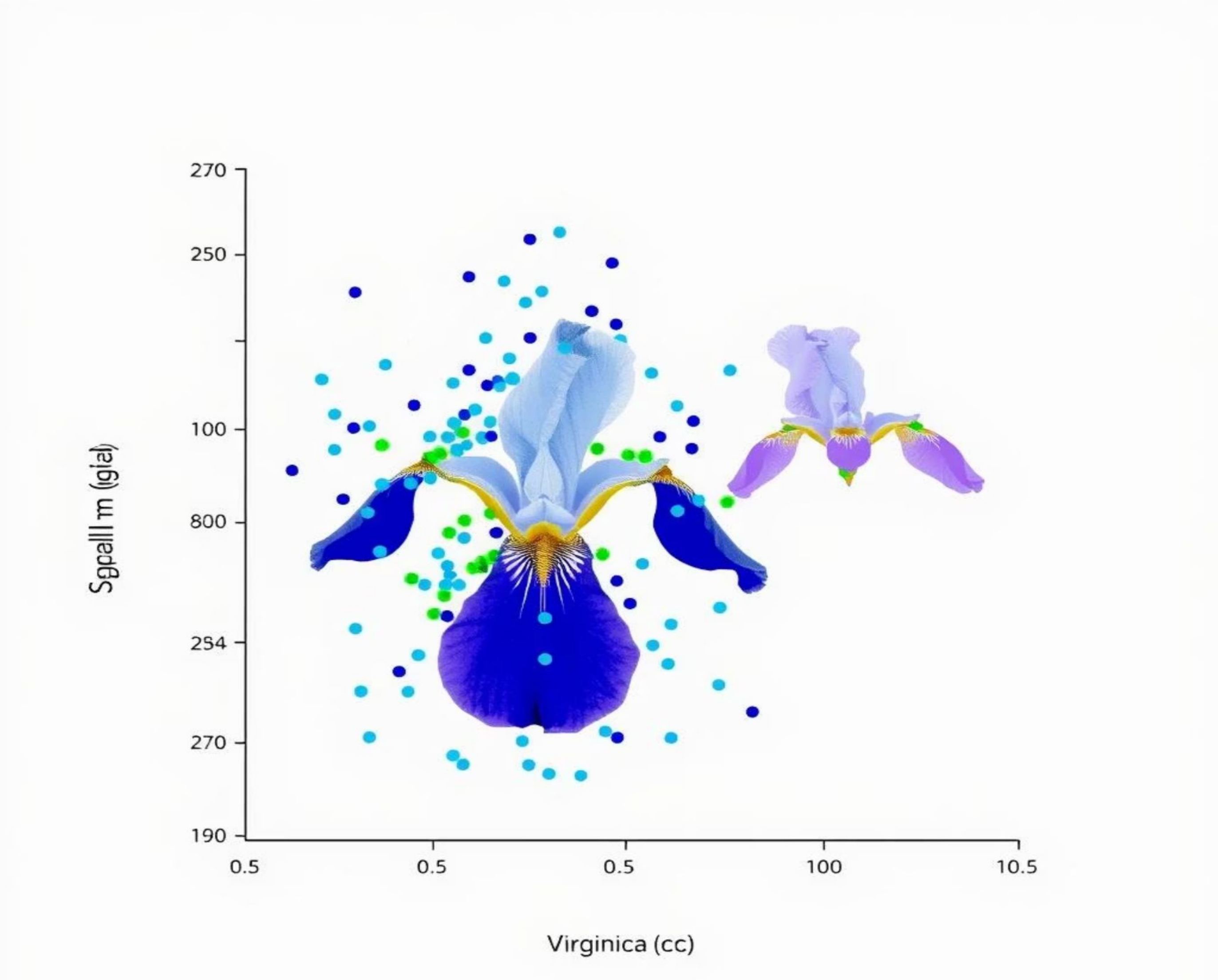
Sepal Width: The width of the sepal.

Petal Length: The length of the petal.

Petal Width: The width of the petal.

The Objective

The primary goal when working with the Iris dataset is to develop an ML model that can accurately predict the species of an Iris flower based solely on these four measurements.



Each data point corresponds to one of three Iris species: Setosa, Versicolour, or Virginica, making it an ideal candidate for multi-class classification.

Practical Application: Code Walkthrough

Let's briefly examine a Python code snippet using the popular Scikit-learn library to implement a Logistic Regression model on a simple dataset and interpret its output.

Logistic Regression Code Snippet

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix
import pandas as pd # Assuming Iris-like data is loaded as a
DataFrame

# Load dataset (example placeholder)
# X, y = load_iris(return_X_y=True)

# Split data
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.3, random_state=42
)
```

```
# Initialise and train model
model = LogisticRegression(max_iter=200)
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate
accuracy = accuracy_score(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)
```

Practical Application: Code Walkthrough (Continued)

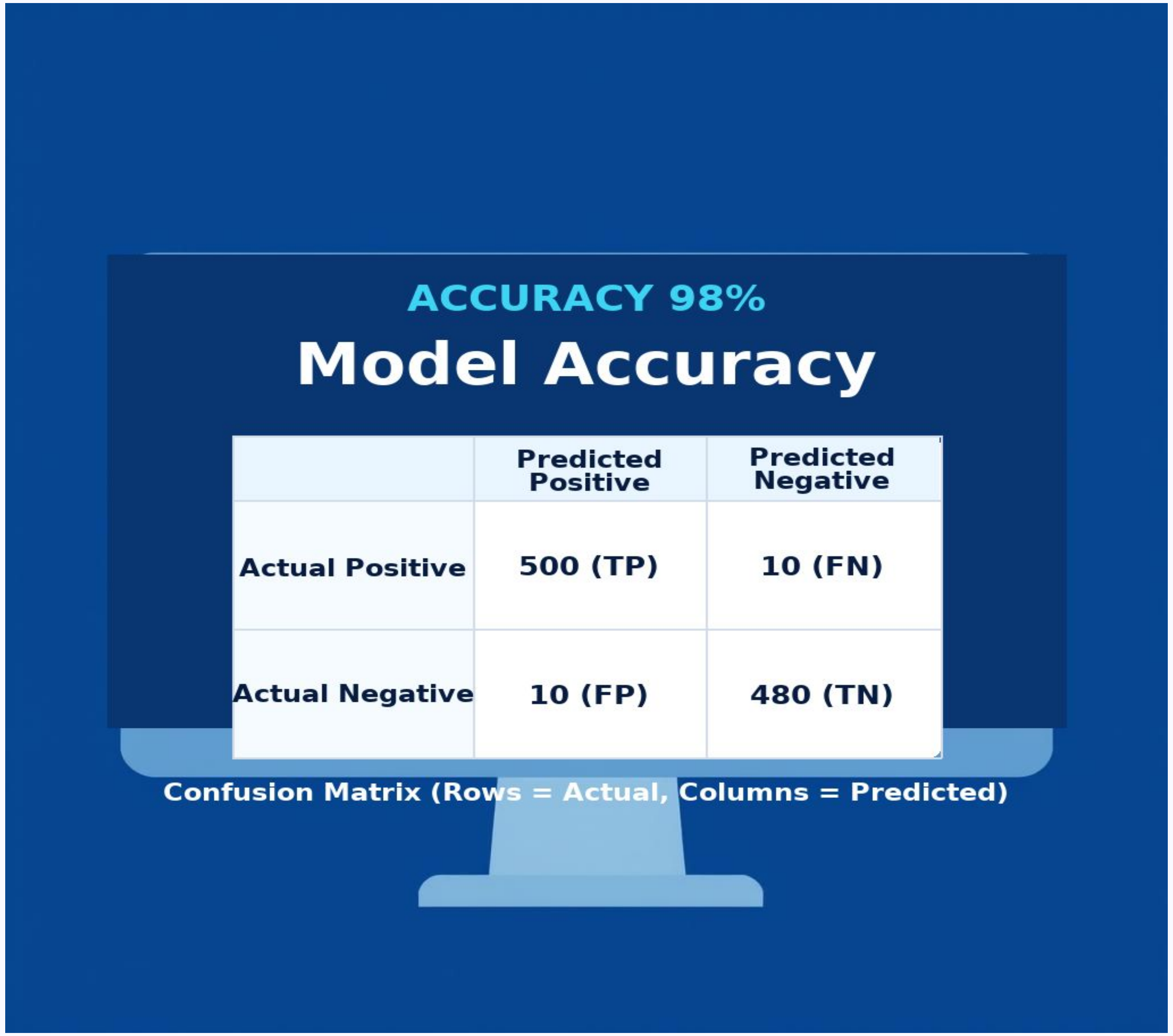
Let's briefly examine a Python code snippet using the popular Scikit-learn library to implement a Logistic Regression model on a simple dataset and interpret its output.

Interpreting the Output

After training our Logistic Regression model, we evaluate its performance using key metrics:

Accuracy: Represents the proportion of correctly classified instances. An accuracy of 0.98 means 98% of predictions were correct.

Confusion Matrix: Provides a detailed breakdown of correct and incorrect classifications. It shows true positives, true negatives, false positives, and false negatives, helping to understand where the model makes errors.



Practical Application: Code Walkthrough (Continued)

Understanding Accuracy and Confusion Matrix

1. Accuracy

- Formula: $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
- Measures the **overall correctness** of the model.
- Example: With 500 TP, 480 TN, 10 FP, 10 FN → Accuracy = 0.98 (98%).
- Limitation: Can be **misleading with imbalanced datasets** (e.g., predicting all negatives in a rare disease dataset may still give high accuracy).

Practical Application: Code Walkthrough (Continued)

2. Confusion Matrix (2×2 example)

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP): 500	False Negative (FN): 10
Actual Negative	False Positive (FP): 10	True Negative (TN): 480

- **True Positive (TP):** Correctly predicted positive cases.
- **True Negative (TN):** Correctly predicted negative cases.
- **False Positive (FP):** Incorrectly predicted as positive.
- **False Negative (FN):** Missed positives, predicted as negative.

Practical Application: Code Walkthrough (Continued)

3. Why Use Both?

- Accuracy gives a quick snapshot.
- Confusion Matrix reveals **where mistakes happen**.
- Helps calculate other metrics:
  - **Precision (TP ÷ (TP + FP))** → How reliable positive predictions are.
  - **Recall (TP ÷ (TP + FN))** → How well positives are captured.
  - **F1 Score** → Balance between precision and recall.

📌 **Takeaway:**
Always look **beyond accuracy** — use confusion matrix and derived metrics to fully understand model performance, especially with **imbalanced data**.

Charting Your Path: Career in Machine Learning

The field of Machine Learning offers diverse and rewarding career opportunities across various industries, constantly evolving and demanding a blend of technical expertise and problem-solving skills.

Essential Skillset

Python: The lingua franca of ML, crucial for scripting and framework interaction.

Numpy & Pandas: Fundamental libraries for numerical computation and data manipulation.

Scikit-learn: A comprehensive library for classic ML algorithms.

TensorFlow/PyTorch: Deep Learning frameworks for advanced model development.

Recommended Roadmap

Basics: Strong foundation in mathematics, statistics, and programming.

Machine Learning: Understanding core algorithms and concepts.

Deep Learning: Delving into neural networks and complex architectures.

NLP/CV: Specialising in Natural Language Processing or Computer Vision.

Projects: Building a robust portfolio with practical applications.

Key Industries

Healthcare: Drug discovery, diagnostics, personalised medicine.

Finance: Algorithmic trading, risk assessment, fraud prevention.

Retail: Recommendation systems, inventory optimisation, customer behaviour analysis.

Autonomous Systems: Self-driving cars, drones, robotics.

Education: Personalised learning platforms, content curation.

Wrapping Up & Your Questions Answered

We've covered a vast landscape of Machine Learning today. Let's consolidate our learning and open the floor for discussion.

Key Takeaways

- ML allows machines to **learn from data** without explicit programming.
- It’s distinct from AI (broader field) and DL (subset of ML).
- ML powers **numerous real-world applications**, from spam filters to self-driving cars.
- Three main types: **Supervised, Unsupervised, and Reinforcement Learning**.
- The **ML pipeline** is a structured process for model development.
- A career in ML requires continuous learning and practical project experience.

Food for Thought: "Where do you see Machine Learning impacting your career path, or a field you're passionate about?"

Wrapping Up & Your Questions Answered

We've covered a vast landscape of Machine Learning today. Let's consolidate our learning and open the floor for discussion.

Resources for Self-Learning

Online Courses: Coursera, edX, DataCamp, fast.ai, Udacity.

Books: "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow", "Deep Learning Book".

Communities: Kaggle, GitHub, Stack Overflow, local meetups.

Blogs: Towards Data Science, Medium articles from experts.

Thank You!

We appreciate your engagement and participation today. Your journey into the world of AI/ML has just begun!

Join us next time for **Class 03: Thesis Roadmap on 8th September 9 PM**