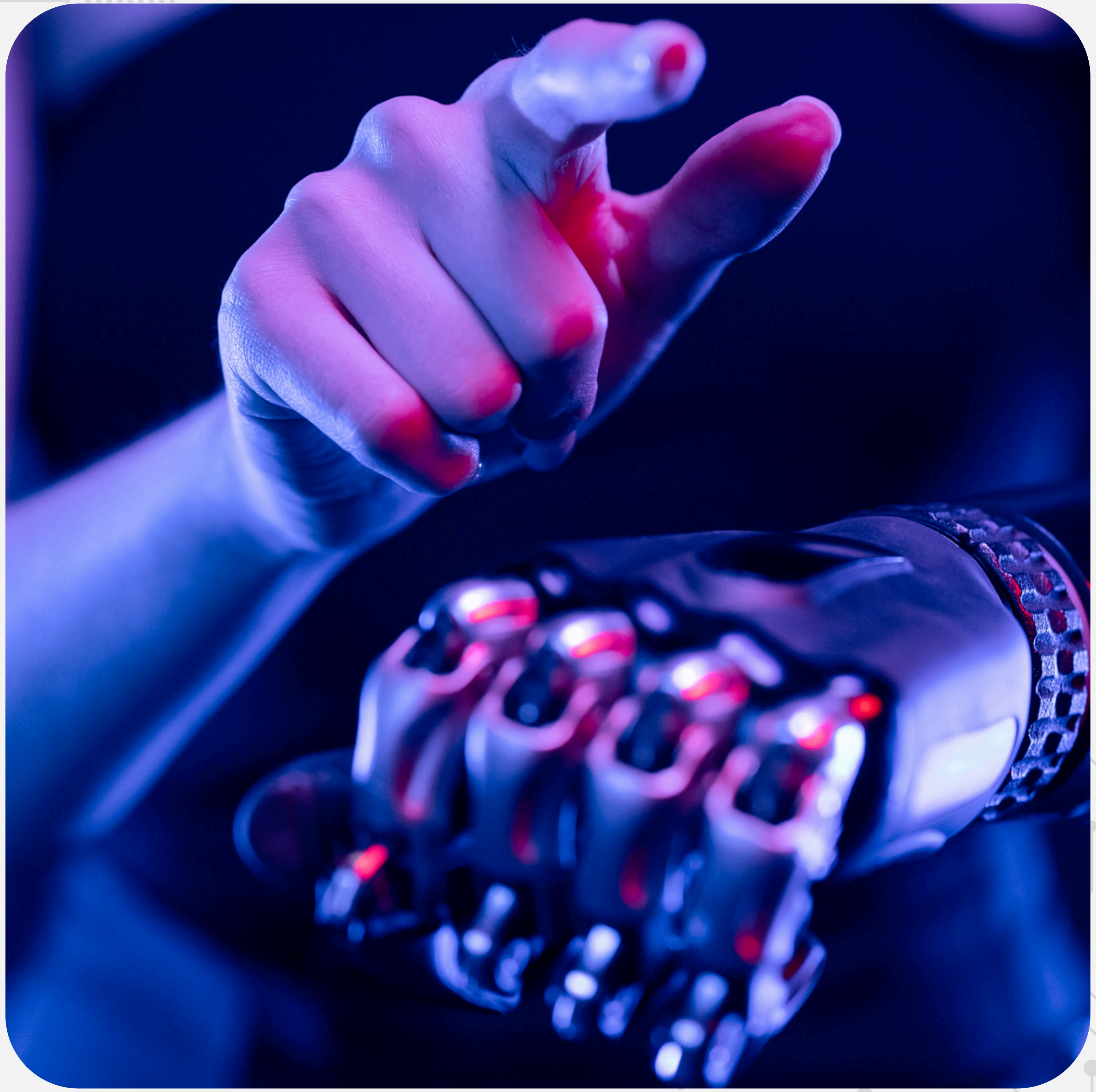
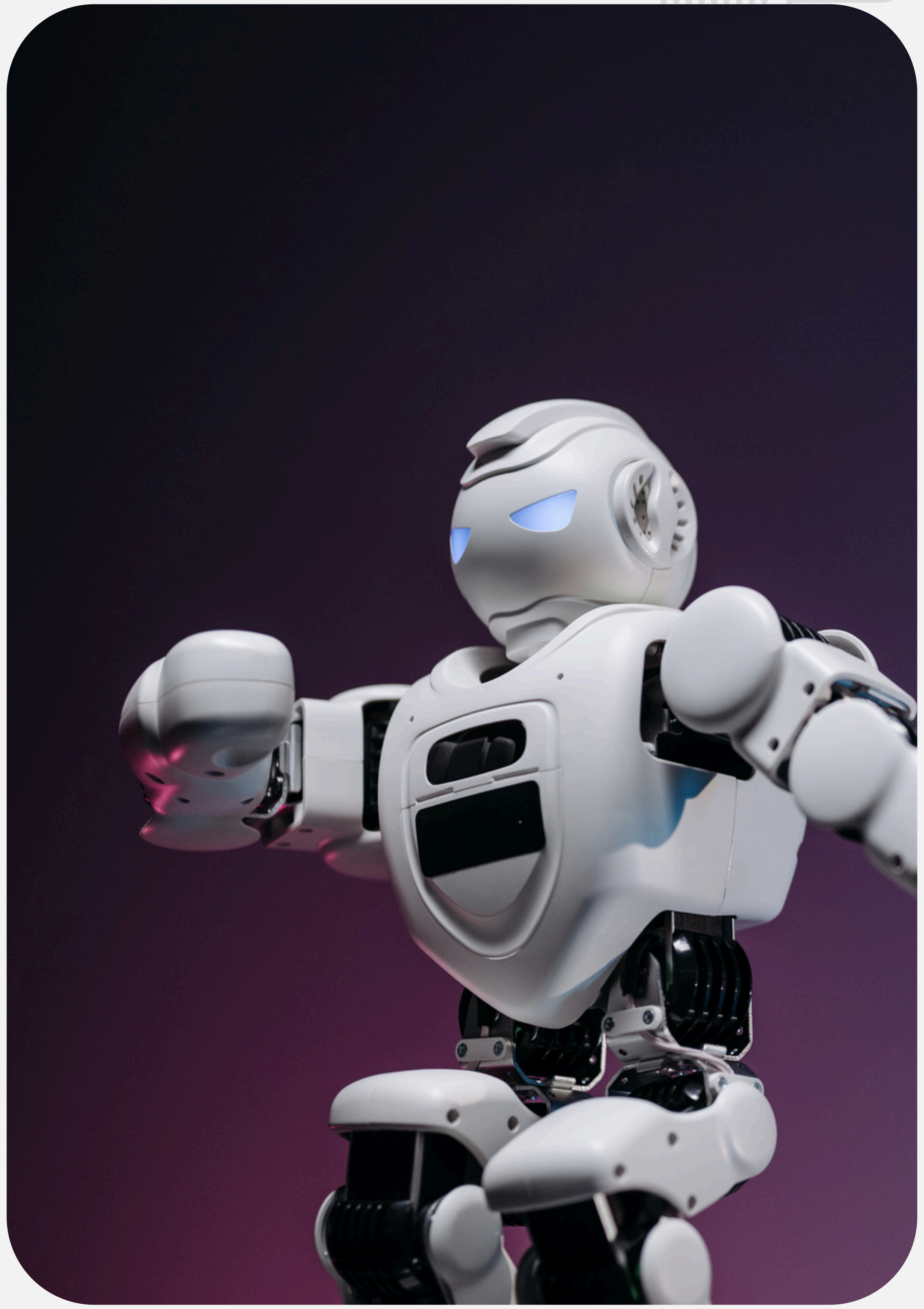
AI & Data Science

Machine Learning





What is Machine Learning?

Machine learning is a branch of Artificial Intelligence that focuses on developing models and algorithms that let computers learn from data without being explicitly programmed for every task. In simple words, ML teaches the systems to think and understand like humans by learning from the data.

Machine Learning Use Cases

Image Recognition

Used in facial recognition, self-driving cars and medical imaging.

Natural Language Processing (NLP)

Used for chatbots, translation and sentiment analysis

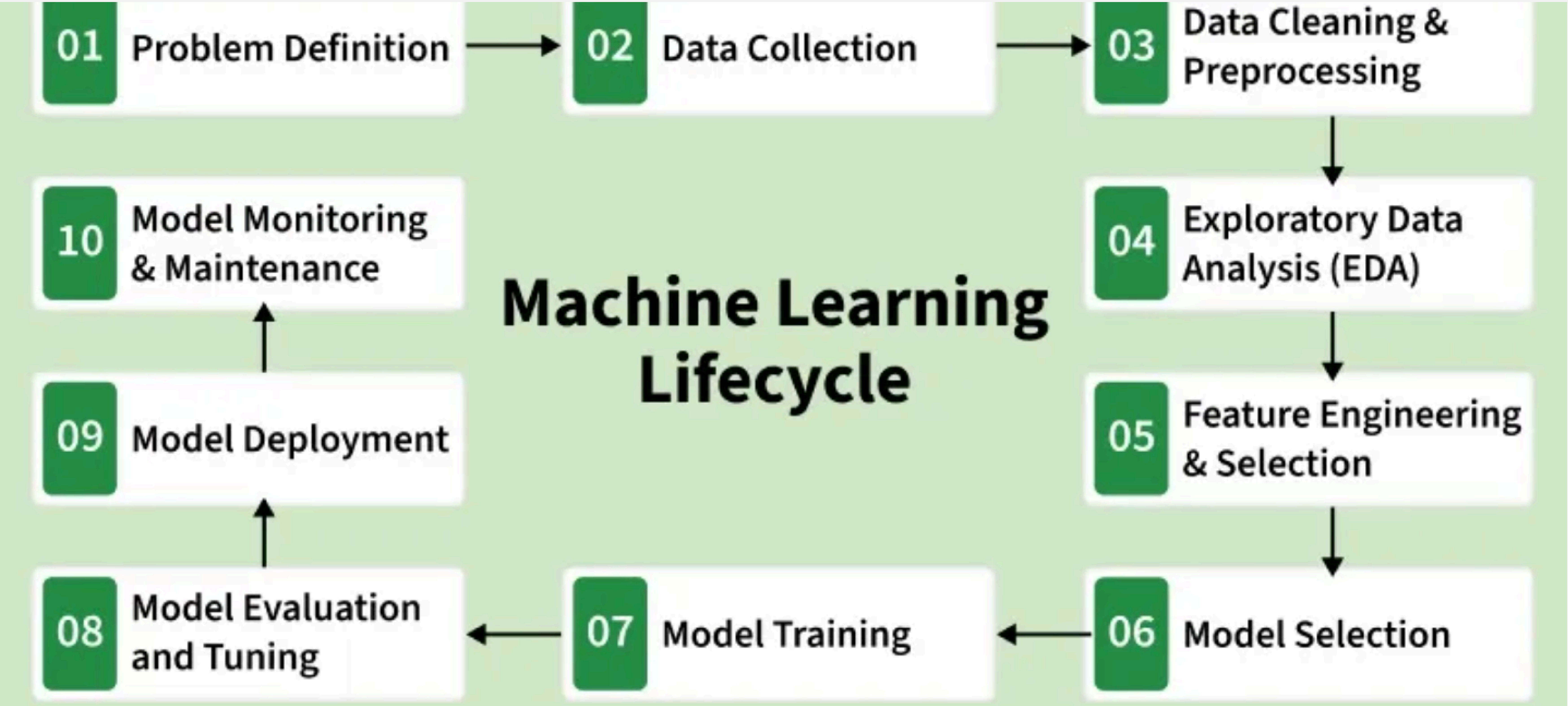
Recommendation Systems

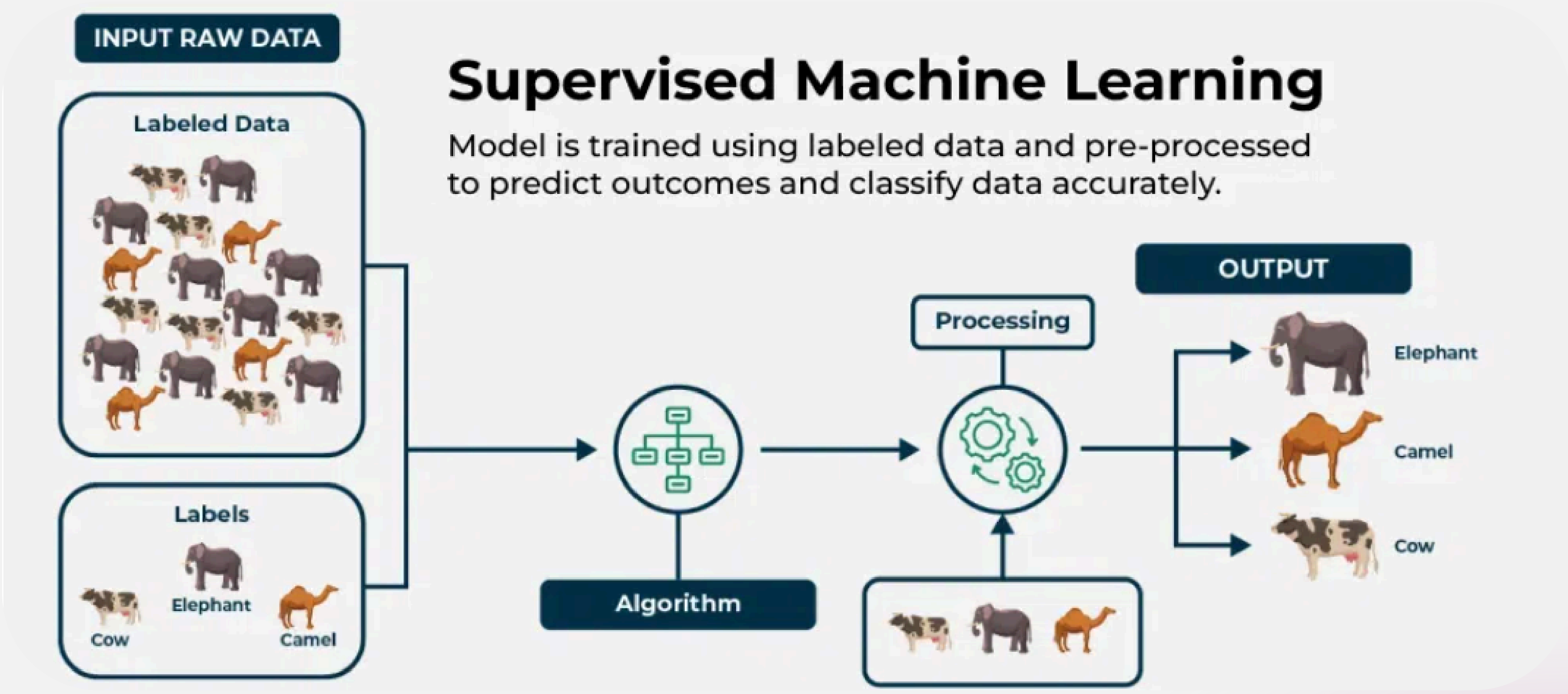
Netflix, Amazon, Spotify.

Predictive Maintenance

Detects machine issues before they happen.

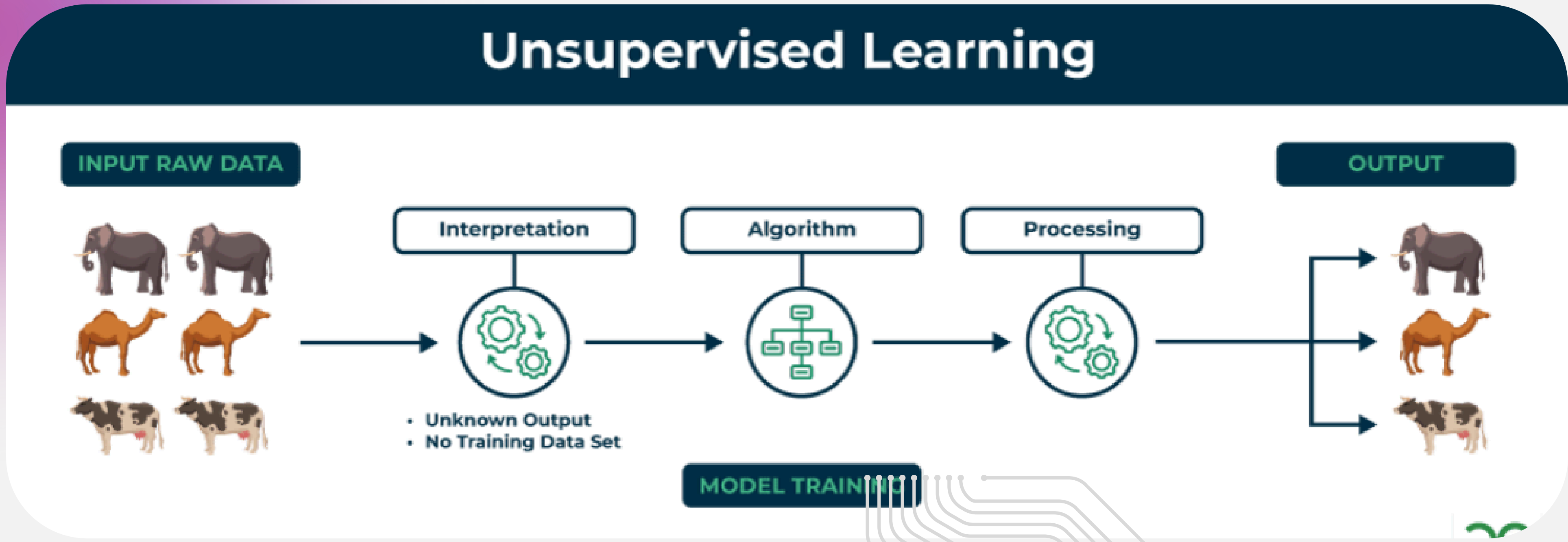
Machine Learning Lifecycle is a structured process that defines how machine learning (ML) models are developed, deployed and maintained. It consists of a series of steps that .ensure the model is accurate, reliable and scalable





- **Supervised Learning**: Trains models on labeled data to predict or classify new, unseen data.
- Supervised learning algorithms are generally categorized into two main types:
- Classification - where the goal is to predict discrete labels or categories
- Regression - where the aim is to predict continuous numerical values.

- **Unsupervised Learning**: Finds patterns or groups in unlabeled data, like clustering or dimensionality reduction.
- Unsupervised learning are again divided into three main categories based on their purpose:
- Clustering
- Association Rule Mining
- Dimensionality Reduction.



supervised learning algorithms

01 Linear Regression

This is one of the simplest ways to predict numbers using a straight line. It helps find the relationship between input and output,

02 Logistic Regression

Used when the output is a "yes or no" type answer. It helps in predicting categories like pass/fail or spam/not spam.

03 Decision Trees

A model that makes decisions by asking a series of simple questions, like a flowchart. Easy to understand and use.

Unsupervised learning algorithms

01 Clustering

Clustering algorithms group data points into clusters based on their similarities or differences.

02 Dimensionality Reduction

Dimensionality reduction is used to simplify datasets by reducing the number of features while retaining the most important information.

03 Association Rule

Find patterns between items in large datasets typically in market basket analysis.

Regression Algorithms

Goal: To predict a continuous value. Output Type: Real-valued numbers (e.g., a price, a temperature, an age). Task: Finding a "best-fit" line or curve to model the relationship between variables.

Algorithm	Description	Common Use Cases
Linear Regression	A simple algorithm that models a **linear relationship** between one or more input features (X) and a continuous output variable (Y) as $Y = aX + b$.	House price prediction, sales forecasting.
Polynomial Regression	Models a **non-linear relationship** by fitting a polynomial equation to the data points.	Modeling growth rates, predicting biological processes.
Ridge and Lasso Regression	**Regularization** techniques (modified versions of Linear Regression) used to prevent overfitting and select the most important features.	Financial modeling, high-dimensional data prediction.
Decision Tree Regression	Uses the same tree structure as classification, but the leaves contain a **continuous value** (the average of the training data in that leaf) instead of a class label.	Predicting staff turnover time, estimating project completion time.

Linear Regression in Machine learn

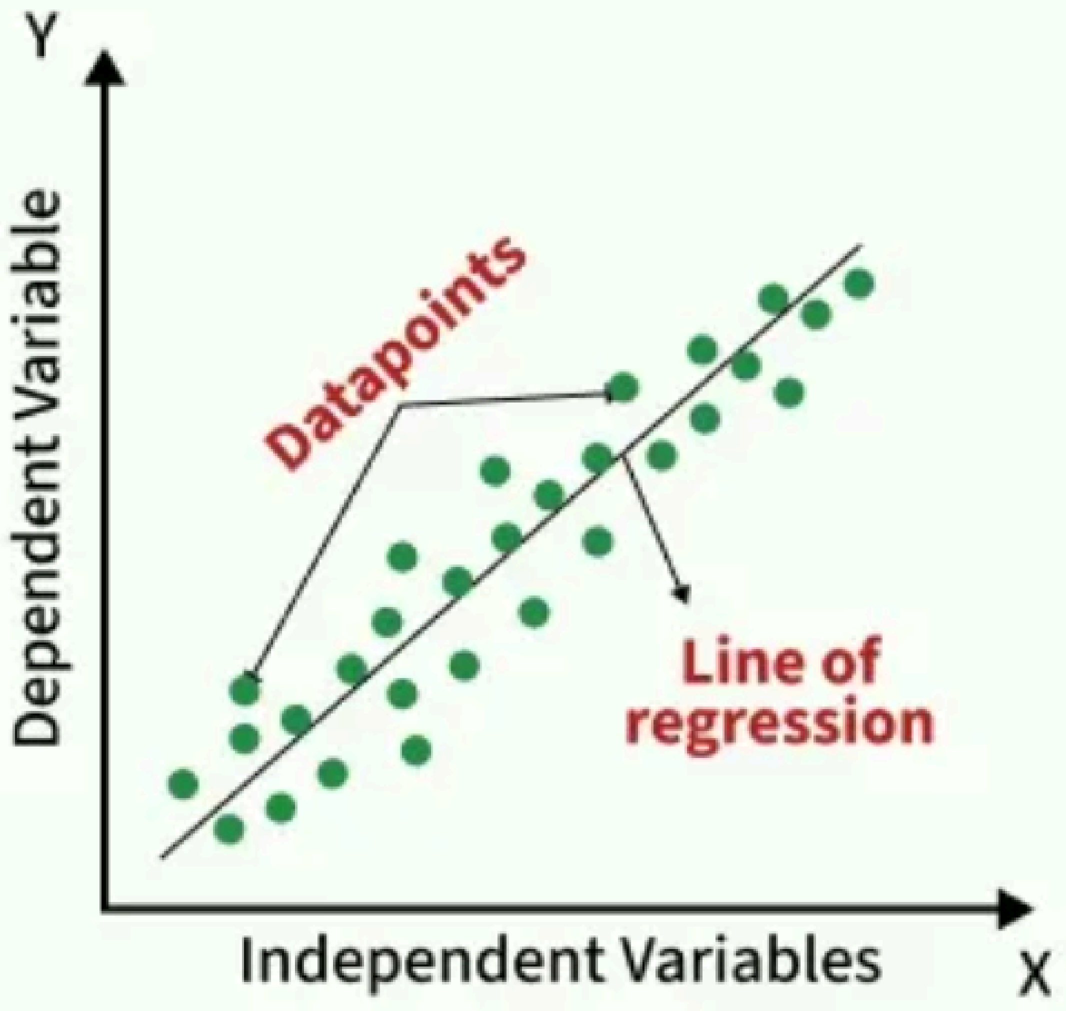
Introduction to Linear Regression

- Simple model for predicting continuous values.
- It finds the relationship between input features and output.
- It is used in forecasting and predicting salaries based on experience.

Linear Regression in Machine learn

How Does Linear Regression Work?

- Finds the best-fitting line by minimizing prediction errors (least squares method)
- It calculates coefficients for variables that minimize the error in predictions.



Linear Regression

Equation of the Best-Fit Line

$$y=mx+b$$

- y is the predicted value (dependent variable)
- x is the input (independent variable)
- m is the slope of the line (how much y changes when x changes)
- b is the intercept (the value of y when x = 0)

Polynomial Regression

- Definition: Extends linear regression by including higher-order terms of the predictor variable.
- Equation:
- $y=b_0+b_1x+b_2x^2+b_3x^3+\dots +b_nx^n$

Classification Algorithms

Goal: To predict a discrete label or category. Output Type: Categorical (e.g., A or B, True or False, red/green/blue).

Task: Drawing a decision boundary to separate data points into classes.

Algorithm	Description	Common Use Cases
Logistic Regression	Despite the name, it's a classification algorithm that estimates the **probability** of an instance belonging to a specific class (often binary).	Spam detection, predicting if a customer will click an ad (Click/No-Click).
Decision Trees	A model that makes predictions by following a series of **if-then-else** decision rules inferred from the data features.	Credit risk assessment, classifying types of cells.
Random Forest	An **ensemble** method that builds multiple Decision Trees and merges their outputs for a more accurate and stable prediction.	Predictive maintenance, image classification.
Support Vector Machines (SVM)	Finds the **optimal hyperplane** that maximizes the margin (distance) between the closest data points of different classes.	Hand-written digit recognition, bioinformatics.
K-Nearest Neighbors (K-NN)	A **non-parametric** method that classifies a new data point based on the majority class of its 'K' nearest data points in the feature space.	Recommendation systems, pattern recognition.
Naive Bayes	A probabilistic classifier based on **Bayes' theorem** with the assumption that features are conditionally independent.	Text classification, sentiment analysis.

Type	Definition	Example Use Cases	Algorithms Commonly Used
Binary Classification	Two categories only	Spam detection, disease diagnosis	Logistic Regression, SVM
Multiclass Classification	More than two categories, one per instance	Digit recognition, sentiment analysis	Softmax Regression, Random Forest
Multi-Label Classification	Multiple categories per instance	Movie tagging, image tagging	Neural Networks, kNN extensions
Imbalanced Classification	Uneven class distribution	Fraud detection, rare diseases	SMOTE, cost-sensitive learning

characteristics of Classification Models

Characteristic	Meaning	Example Use Case
Class Separation	Distinguishing categories clearly	Spam vs. non-spam emails
Decision Boundaries	Line/surface dividing classes	SVM hyperplane for image recognition
Sensitivity to Data Quality	Reliance on clean, representative data	Medical diagnosis with patient records
Handling Imbalanced Data	Managing uneven class distribution	Fraud detection, rare disease
Interpretability	Human understanding of model decisions	Loan approval, healthcare models

Clustering Algorithms

- Definition: Unsupervised learning method that groups data points into clusters based on similarity.
- Examples:
  - K-Means → partitions data into K clusters.
  - Hierarchical Clustering → builds a tree of clusters.

Algorithm	**Requires k?**	**Handles Noise**	**Cluster Shape**	**Best For**
K-Means	Yes	No	Spherical	Large, simple datasets
Hierarchical	No	Limited	Nested clusters	Small datasets, visualization
DBSCAN	No	Yes	Arbitrary	Anomaly detection, irregular shapes

Dimensionality Reduction

- Definition: Technique to reduce the number of input variables/features while preserving most of the data's variance.
- Examples:
  - Principal Component Analysis (PCA) → projects data into fewer dimensions.

Technique	Type	How It Works	Best Use Cases
Principal Component Analysis (PCA)	Linear, unsupervised	Finds new axes (principal components) that maximize variance in the data [1]	Image compression, text feature reduction
Linear Discriminant Analysis (LDA)	Linear, supervised	Projects data to maximize class separability [1]	Classification tasks (face recognition, handwriting recognition)
t-SNE (t-Distributed Stochastic Neighbor Embedding)	Nonlinear, unsupervised	Preserves local similarities by mapping high-dimensional data into 2D/3D [1]	Visualizing embeddings, clustering
Independent Component Analysis (ICA)	Linear, unsupervised	Separates signals into statistically independent components [1]	Signal processing, blind source separation
Singular Value Decomposition (SVD)	Linear, unsupervised	Factorizes a matrix into singular values and vectors [2]	Latent Semantic Analysis (topic modeling in NLP)
Feature Selection	Linear/ nonlinear	Selects a subset of relevant features using filters, wrappers, or	Preprocessing for regression/ classification

Gradient Descent

Gradient Descent is an optimization algorithm used in machine learning to minimize a cost (loss) function by iteratively adjusting model parameters in the direction of steepest descent of the gradient.

Type	How It Works	Pros	Cons	Best Use Cases
Batch Gradient Descent	Uses the entire dataset to compute gradients before updating parameters	Stable convergence, global view of loss	Slow for large datasets, memory intensive	Small to medium datasets
Stochastic Gradient Descent (SGD)	Updates parameters using one training example at a time	Fast updates, works well for large datasets	Noisy updates, may overshoot minima	Online learning, deep learning
Mini-Batch Gradient Descent	Uses a subset (batch) of data for each update	Balance between speed and stability, efficient on GPUs	Requires tuning batch size	Deep learning, large datasets

Aspect	Overfitting	Underfitting
Definition	Model learns training data too well, including noise and irrelevant details.	Model is too simple and fails to capture underlying trends.
Training Performance	Very high accuracy (low error).	Low accuracy (high error).
Test/ Validation Performance	Poor generalization; high error on unseen data.	Poor generalization; also high error on unseen data.
Bias & Variance	Low bias, high variance.	High bias, low variance.
Model Complexity	Too complex (e.g., deep trees, high-degree polynomials).	Too simple (e.g., linear regression on non-linear data).
Analogy	Student memorizes answers but struggles with new questions.	Student barely studies and cannot answer even familiar questions.
Example	A polynomial curve that passes through every training point but fails on new points.	A straight line fitted to data with a clear curve.

Bias:
Bias is the error from simplifying assumptions in the model.
High bias means the model ignores important patterns → underfitting.

Example: Using a straight line to fit curved data.

variance:

- .Variance is the error from sensitivity to training data fluctuations
- .High variance means the model memorizes noise → overfitting
- Example: A high-degree polynomial curve that fits every training point but fails on test .data

<u>Underfit(bias)</u>
High bais
Low model complexing
Poor preformane on both data training and testing
High accuracy
Low varience

Overfit model:(variance)
low bais
high model complexing
high preformane on both data training and testing
low accuracy
High varience

Support Vector Machines

Support Vector Machines (SVMs) are supervised learning algorithms that classify or regress .data by finding the optimal hyperplane that maximizes the margin between classes

Hyperplane: The decision boundary separating classes. In 2D, it’s a line; in higher dimensions, it’s a plane.

Support Vectors: Data points closest to the hyperplane. They define the margin and are critical for .classification

Margin: Distance between the hyperplane and support vectors. SVM maximizes this margin for .better generalization

Kernel Trick: Transforms non-linearly separable data into higher dimensions (e.g., polynomial, radial .basis function kernels

:**Hard vs. Soft Margin**

.Hard Margin: Perfect separation, no misclassifications

.Soft Margin: Allows some misclassifications, balancing flexibility and accuracy

Regularization (C parameter): Controls trade-off between maximizing margin and minimizing .classification errors

Dimensionality Reduction

Dimensionality Reduction means simplifying a dataset by reducing the number of input variables (features) while keeping the most important information intact.

Principal Component Analysis (PCA)
Goal: Find new axes (principal components) that capture the maximum variance in the data

- Intuition: PCA rotates the coordinate system to align with directions of maximum variance.

Singular Value Decomposition (SVD)

- Goal: Factorize a matrix into simpler components to capture its structure.
- Intuition: SVD compresses data by keeping only the largest singular values (most important directions).

Linear Discriminant Analysis (LDA)

- Goal: Unlike PCA (unsupervised), LDA is supervised and aims to maximize class separability.
- Intuition: LDA finds axes that best separate categories

.

Technique	**Supervised?**	**Preserves**	**Best For**
PCA	No	Variance	Compression, visualization
SVD	No	Matrix structure	NLP, topic modeling
LDA	Yes	Class separability	Classification tasks

Decsion tree

- Definition: A non-parametric supervised learning method used for classification (categorical outcomes) and regression (continuous outcomes).
- Root Node: Represents the entire dataset.
- Branches: Paths based on feature conditions.
- Internal Nodes: Decision points where data is split.
- Leaf Nodes: Final predictions or classifications

How it works

Each internal node represents a feature

Each branch represents a decision rule

Each leaf node represents a final prediction

Advantages

Very easy to understand and visualize

Fast to train

Works well with small datasets

Disadvantages

Highly prone to overfitting

Sensitive to small changes in the data

Not very robust to outliers

Random Forest

Random Forest is a powerful ensemble machine learning algorithm that builds many decision trees and combines their outputs. For classification, it uses majority voting; for regression, it averages predictions. This approach reduces overfitting and improves accuracy compared to a single decision tree.

How it works

Creates multiple bootstrap samples from the dataset
Trains a separate decision tree on each sample
Uses a random subset of features at each split
Combines all tree predictions:
Majority vote for classification
Average for regression

Advantages

Much lower risk of overfitting
More accurate and stable than a single tree
Robust to noise and outliers

Disadvantages

Harder to interpret
Slower to train and predict
Requires more computational resources

Ensemble Learning

Definition: A technique that aggregates predictions from multiple models (weak or strong learners).

Method	How It Works	Strengths	Weaknesses	Examples
Bagging (Bootstrap Aggregating)	Trains models independently on random subsets of data, then aggregates predictions (majority vote or averaging).	Reduces variance, prevents overfitting.	Requires many models, computationally heavy.	Random Forest
Boosting	Models are trained sequentially; each new model focuses on correcting errors of the previous one.	Reduces bias, achieves high accuracy.	Sensitive to noise, risk of overfitting.	AdaBoost, Gradient Boosting, XGBoost
Stacking (Stacked Generalization)	Combines predictions of multiple models using a meta-model that learns how to best blend them.	Flexible, can combine different algorithms.	Complex, requires careful tuning.	Stacked ensembles in Kaggle competitions

Thank You!