What is Supervised Learning?

Supervised learning is a type of machine learning where the model learns from **labeled data**.

- **Labeled data** means each training example has both an **input** (features) and a known **output** (label).
- The model tries to **find patterns** that map inputs to correct outputs.
- Once trained, the model can predict the output for **new, unseen data**.
- The main idea is similar to a teacher supervising a student , the correct answers are known and the model learns by comparing its predictions with the true answers and adjusting itself.

Common tasks:

- **Classification:** Predicting categories.
- **Regression:** Predicting continuous values.

Examples:

- Handwritten digit recognition
- Weather forecasting based on past climate data
- Credit card fraud detection

What is Unsupervised Learning?

Unsupervised learning is a type of machine learning where the model works with **unlabeled data**.

- The data only has **inputs**, with no correct outputs given.
- The goal is to **discover hidden structures**, patterns, or relationships in the data.
- There is no teacher the model organizes or simplifies the data by itself.

Common tasks:

- **Clustering:** Grouping similar data points together.
- **Dimensionality Reduction:** Reducing the number of features while keeping important information. (e.g., PCA : Principal Component Analysis)
- **Association Rules:** Finding relationships between variables in large datasets. (e.g., market basket analysis)

Examples:

- Organizing news articles into topics
- Finding groups of genes with similar expression patterns
- Anomaly detection

Dataset: Credit card fraud detection

Columns:

- `step` — Time step of the transaction
- `type` — Transaction type (PAYMENT, TRANSFER, CASH_OUT, etc.)
- `amount` — Transaction amount
- `nameOrig` — Origin account name
- `oldbalanceOrg` — Original balance of sender
- `newbalanceOrig` — New balance of sender after transaction
- `nameDest` — Destination account name
- `oldbalanceDest` — Original balance of receiver
- `newbalanceDest` — New balance of receiver after transaction
- `isFraud` — Whether the transaction is fraudulent (0 or 1)
- `isFlaggedFraud` — Whether the system flagged it as suspicious

Suitable for Supervised Learning?

Absolutely yes.

- We have a clear **label**: `isFraud` (0 or 1)
- We can train a **classification model** to predict whether a transaction is fraud or not, using the other features.
- This is a **classic fraud detection problem** in supervised learning.

Suitable for Unsupervised Learning?

Also yes.

- We can just **ignore the labels** (`isFraud` and `isFlaggedFraud`) and apply **clustering** to discover hidden patterns.
- For example:
    - Find clusters of unusual transactions that stand out from normal ones.
    - Detect anomalies some fraud detection systems combine clustering and outlier detection to find suspicious cases the supervised model may miss.