Machine Learning: Features and Feature Engineering

What are Features?

Features are the attributes or properties that help distinguish one object from another.

For example, suppose there are two people named Arif in your class — one wears glasses, the other doesn't. If you want to identify which is which, you might say, **"the one who wears glasses" or "the one without glasses"**. Here, wearing glasses is a feature.

How do Computers Learn?

When we want a computer to recognize something, we need to **train it with data**.
For instance, if we want the computer to identify a giraffe:

- We can feed it many labeled photos of giraffes and non-giraffes.
- Each photo is described by its features (like neck length, tail length, or whether it has horns).

Alternatively, we can enable the computer to extract its own features directly from the images.

Supervised vs Unsupervised Machine Learning:

Supervised Machine Learning:
We provide labeled data — which means we tell the algorithm which photos are giraffes and which are not.

Unsupervised Machine Learning:
We do not provide any label. The algorithm finds patterns by itself.
(In this case, the label column is missing in the table).

How This Works:

The algorithm first:

- Detects the neck, the tail, or the horns in a new picture.
- Measures their lengths or checks their presence.
- Compares these with its training data.
- Prints "Yes" if it's a giraffe, "No" if not.

What's a Response Variable?

The response variable is **the output the algorithm produces** (like 0 or 1).

- If output == 1 → Giraffe
- If output == 0 → Not Giraffe

The label is what this output means.
(In this case: 1 = Giraffe, 0 = Not Giraffe)

What is Feature Engineering?

Feature engineering is a crucial process in Machine Learning and Data Analysis.
It involves **creating or selecting helpful features** from raw data to aid the algorithm in making accurate predictions.

Main Processes in Feature Engineering:

Feature Selection: Choosing the most relevant or important features.

Feature Transformation: Modifying or creating new features from the existing ones.

Feature Creation: Generating new features from raw data.

Data Scaling and Normalization: Transforming the range of values to aid algorithm performance.

Example:

Let's say we want to predict the electricity consumption of a building.
We have the following data:

Temp (°C)	Humidity (%)	Time (Hours)	Electricity Consumption (Units)
30	70	13	150
35	65	14	200
25	80	20	100

Feature Selection :

- **Temperature** and **humidity** might be more important.
- Time might be less relevant.

Feature Transformation :

We can combine temperature and humidity to create a new feature, **Weather Impact**.

Feature Creation :

We can create a new feature called **High Temperature** (Yes/No) if temperature > 32°C.

Temp (°C)	Humidity (%)	High Temp	Electricity Consumption
30	70	No	150
35	65	Yes	200
25	80	No	100

Data Scaling and Normalization :

We can normalize the values to range from 0 to 1.

Importantly:

Feature engineering is both **an art and a science**.
Properly performed, it can significantly improve your algorithm’s performance.
Incorrect or poor feature engineering can reduce its accuracy.
This is a key step in developing a successful Machine Learning Model.