* Encoding Categorical Data:- (Ordinal Encoding)

(Day - 26)

"Encoding categorical Data means converting categories (text/labels/words) into numbers so that machine learning algo. can understand and use them.

→ Apda dataset ma kai-kai features string/text ma hoy che (jaise ke "Color", "City", "Gender"). Computer ane ML models ne text samajhatu nathi, tene number format joie.

⇒ Types of Encoding :-

(1) Label Encoding :- (Nominal)

"Each category is assigned a unique number."

→ Eg :- Color = {Red, Green, Blue}.

Red → 0, Green → 1, Blue → 2

→ Pro :- Simple to use :-

→ Problem :- Model may assume order/priority bet^n numbers (like 2 > 1 > 0), even if no real order exists.

→ This transformer should be used to encode target values, i.e. y, and not input x.

(2) One-Hot Encoding :- (Nominal)

"Each Category is converted into a new column with binary values (0 or 1).

→ Eg :- Color = {Red, Green, Blue}

Red → [1,0,0], Green → [0,1,0], Blue [0,0,1]

→ Pro :- No false order problem.

→ Con :- Increases the no. of columns. (high dimensional data).

(3) Ordinal Encoding:-

"Categories are replaced with numbers based on their rank/order.

Eg:- Size = {Small, Medium, Large}.
— Small → 1, Medium → 2, Large → 3.

→ Pro :- Maintains order of categories.

→ Cons:- Should only be used when data has a natural order.