

Categorical Encoding

- **Categorical Data को Numbers में बदलना पड़ता है** इसी process को कहते हैं **Categorical Encoding**
- Machine Learning में डेटा दो तरह का होता है:
- 1. Numerical Data (संख्यात्मक) Age → 25, Salary → 30000
- 2. Categorical Data (श्रेणीबद्ध) Gender → Male / Female
- City → Delhi / Mumbai / Jaipur
- Color → Red / Blue / Green
- Machine Learning models सीधे **text (शब्द)** नहीं समझते।
उन्हें सिर्फ **numbers** समझ आते हैं।
- Machine Learning algorithm बोलेगा: → “ये Delhi क्या है? मुझे तो नंबर चाहिए!”
- Delhi → 1,
- Mumbai → 2



Type of Categorical Encoding

- Nominal Data कोई order नहीं होता Color → Red, Blue, Green, City → Delhi, Mumbai,
- Ordinal Data order होता है Size → Small < Medium < Large, Education → 10th < 12th < Graduate
- **Categorical Encoding के Types**
 - Type 1: Label Encoding
 - Type 3: Ordinal Encoding
 - Type 3: One Hot Encoding (सबसे Popular)
 - Type 4: Binary Encoding
 - Type 5: Target Encoding
- गलत encoding → गलत prediction,
- One Hot Encoding सबसे safe option है beginners के लिए

1. Label Encoding

- **Label Encoding** एक technique है जिसमें हर **category (text)** को एक **unique integer number** दे दिया जाता है।
- City = ['Delhi', 'Mumbai', 'Delhi'] → Encoded = [0, 1, 0]
- Dependent (Output / y) के लिए बहुत ज़्यादा use होता है
- Label Encoding use for 1D/Single Column k liye
- Label Encoding हमेशा 1D होता है
- `from sklearn.preprocessing import LabelEncoder`

2. Ordinal Encoding

- **Ordinal Encoding** एक encoding technique है जिसमें **categorical data** को **numbers** में बदला जाता है
- Ordinal Data = ऐसा categorical data जिसमें natural order होता है
- Examples:
 - Size \rightarrow Small $<$ Medium $<$ Large
 - Education \rightarrow 10th $<$ 12th $<$ Graduate $<$ Post-Graduate
 - Rating \rightarrow Poor $<$ Average $<$ Good $<$ Excellent
- अगर order नहीं है \rightarrow Ordinal Encoding use मत करो
- Independent Variable (Input / X) मुख्य रूप से यहीं use होता है
- कब use करें: जब input feature **ordinal** हो जब order meaningful हो
- **Label Encoding use for 2D/Multiple Column k liye**

What

- **Machine Learning** = Data + Learning + Prediction

What

- **Machine Learning** = Data + Learning + Prediction

What

- **Machine Learning** = Data + Learning + Prediction

What

- **Machine Learning** = Data + Learning + Prediction