

FEATURE ENCODING

- Process of converting categorical data into numerical format.
- Important for handling features with string, label or category
- ★ All model require Feature Encoding.

TYPES OF FEATURE & ITS ENCODING

① Nominal Feature: Categories w/o any intrinsic order or ranking.

E.g ⇒ Color: Red, Blue, Green
 Cities: New York, Delhi, Paris
 Countries: INDIA, USA

Encoding Needed: ① One-Hot Encoding
 ② Label Encoding

(II) Ordinal Feature : Categories with a clear, meaningful order or ranking.

E.g => Rating: Poor, Avg, Good, Excellent
 Education: High School, B.E, Master, PHD

Income Bracket: Low, Medium, High

Encoding Needed: Label or Ordinal Encoding

(III) Binary Features : Features with only two possible values

E.g => Gender : Male or Female
 Flag : True or False
 Ans : Yes or No

Encoding Needed : (i) Directly represent by 1 or 0.

(ii) Label Encoding

(IV) High Cardinality Feature : Features with large no. of unique categories

E.g => Product : 1001, 1002, 1003, ...
 ID



Encoding : ① Target Encoding
Needed ② Embeddings (N.N)

⑤ TEXTUAL FEATURES : Textual data

For Example : Tweet, Reviews
: Product Description

Encoding : ① TF-IDF
Need ② Word2Vec
③ Word Embedding
(Glove).