

Encoding Techniques

↓
Regression Problem

→ target = continuous value

$\{f_1, f_2, \underline{f_3}, f_4, f_5\}$ → target (Price of a house)
↓
color ⇒ $\{Red, Blue, Green\}$

↓
Categorical feature

Motivation of
Learning
Encoding
technique

W/o any kind
of preprocessing

ML Model

Encoding

Techniques

PhD > MBA / M.Tech > B.Tech

→ Nominal

One hot
Encoding

(No or such
ranking)

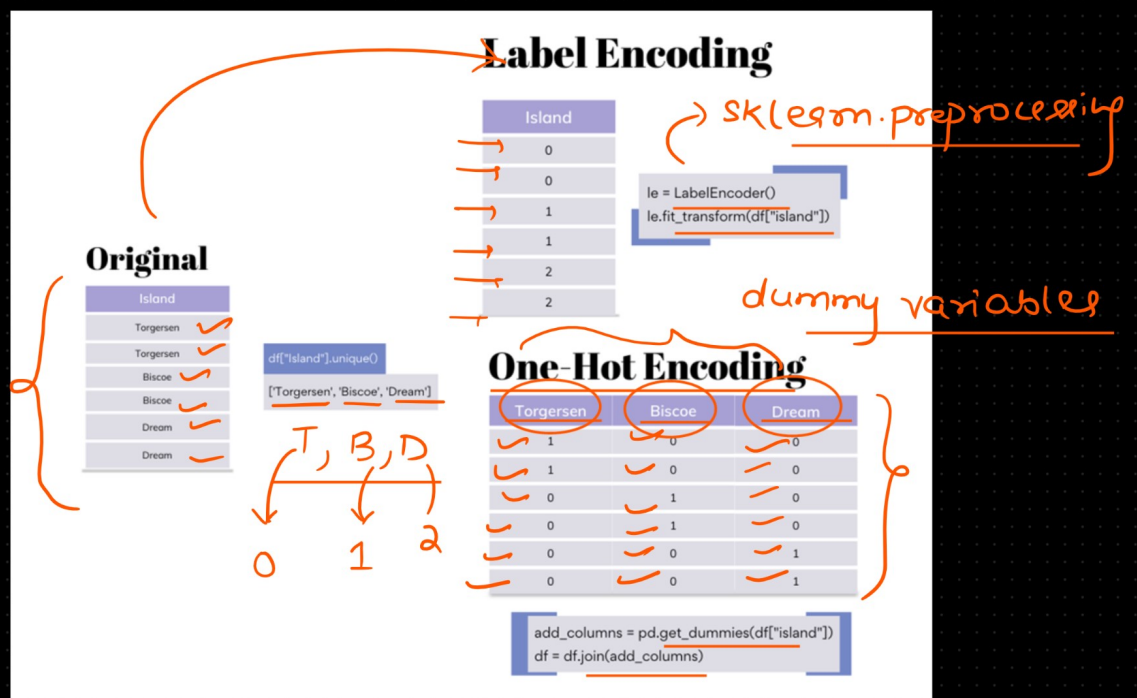
(Ranking)

→ Ordinal

→ Higher
Studies

Label
Encoding

→ B.Tech
→ M.Tech
→ PhD



→ Dummy Variable Trap

Color = Red, Green, Blue, Red, Green, Red

One-hot encoding

multicollinearity

<u>Red</u>	<u>Green</u>	<u>Blue</u>
1	0	0
0	1	0
0	0	1
1	0	0
0	1	0
1	0	0

Why?? → Redundant information

← obviously, its trivial