Label Encoding vs One-Hot Encoding vs Target Encoding

Sample Dataset Demonstration

Sample Dataset

ID	Country	Gender	Purchase
1	USA	Male	200
2	UK	Female	150
3	India	Female	300
4	USA	Female	250

Label Encoding
Country: {USA=2, UK=1, India=0}, Gender: {Male=1, Female=0}

ID	Country	Gender	Purchase
1	2	1	200
2	1	0	150
3	0	0	300
4	2	0	250

One-Hot Encoding

ID	Country_ India	Country_ UK	Country_ USA	Gender_Fe male	Gender_ Male	Purchase
1	0	0	1	0	1	200
2	0	1	0	1	0	150
3	1	0	0	1	0	300
4	0	0	1	1	0	250

Target Encoding
Country averages: India=300, UK=150, USA=225

ID	Country (TE)	Gender	Purchase
1	225	Male	200
2	150	Female	150
3	300	Female	300
4	225	Female	250

Comparison of Encoding Methods

Aspect	Label Encoding	One-Hot Encoding	Target Encoding
Representation	Integers	Binary columns	Mean target values
Implied Order	✅ Yes (misleading)	❌ No	❌ No
Memory Efficiency	✅ Compact	❌ Expands size	✅ Compact
Best For	Tree Models	Linear/Deep Models	High-cardinality features
Risk	Order bias	High dimensionality	Data leakage if careless