The CATegorical Conspiracy

Encoding Cat Food Data

Prompt

How to encode categorical data?

Strategy

We'll use One-Hot Encoding, which makes new binary columns since our data doesn't have any order (e.g., "tuna", "beef", "chicken").

Code & Results

This code adds multiple columns, with each brand receiving its column.

```
data_encoded = pd.get_dummies(data, columns=['company'], drop_first=True)
data_encoded.head(10)
```

	price	weight	star_rating	total_comments	adult	chicken	wet	from_ocean	fish	tuna	...	company_nurture	company_pr
0	189.0	1.2	4.2	8285.0	1.0	0	1.0	1	0	0	...	0	
1	879.0	7.0	4.3	11737.0	1.0	0	1.0	1	1	0	...	0	
2	1900.0	7.0	4.5	8056.0	1.0	0	1.0	1	1	0	...	0	
3	719.0	4.2	4.1	4557.0	1.0	0	1.0	1	1	0	...	0	
4	429.0	2.4	4.0	4522.0	1.0	0	1.0	1	1	0	...	0	
5	121.0	1.0	4.1	3166.0	1.0	0	1.0	0	0	0	...	0	
6	369.0	1.1	4.5	12982.0	0.5	0	1.0	1	1	0	...	0	
7	808.0	3.0	4.5	6133.0	1.0	0	1.0	1	0	1	...	0	
8	387.0	1.2	4.5	8056.0	1.0	0	1.0	1	1	0	...	0	
9	387.0	1.2	4.5	3589.0	1.0	1	1.0	0	0	0	...	0	

10 rows × 48 columns

The number of columns increased from 23 to 48.

💡Tips

- Know why you need encoding
  Machine learning models can't work with text — they need numbers!
- Use **label encoding** for categories with order
  Example: "small" → 0, "medium" → 1, "large" → 2.
  Good when the values have a natural order.
- Use **one-hot encoding** for categories with no order
  Example: "cat", "dog", "fish" becomes three columns — one for each animal.
  Use this when categories are just names.
- Don't encode **numeric-looking** categories as numbers
  Example: zip codes, IDs — they look like numbers but aren't. Treat them as text.
- Watch out for **too many categories**
  If you have 1,000+ unique values (like city names), one-hot encoding can make your dataset explode. You may need to group or simplify.
- Keep track of what each number means
  Save a key or legend so you know what "0", "1", or "2" meant later.
- Be consistent
  Use the same encoding for both training and testing data.

Made by: @katerynakononova / Machine learning: for humans on cats
Watch the short: https://youtube.com/shorts/4pWwKtpgtjU

Website: https://katerynakononova.github.io/meowlearning/

Dive deeper into the world of AI with *Machine Learning: for Humans on Cats* — now on Amazon!
📘👉 https://www.amazon.com/dp/B0CW9SFYXF