



H2O.oidi

"Confidential and property of H2O.ai. All rights reserved"

Cross Validation Target Encoding

Fold	Pay 1	Default Payment	Mean Target Encoding
2	Up To Date	0	0
3	Up To Date	0	0
2	Missed 1 Mo	1	0.5
3	Missed 1 Mo	0	0.5

Fold	Pay 1	Default Payment	CV Target Encoding
1	Up To Date	0	0
1	Missed 1 Mo	0	0.5
1	Missed 5 Mo	1	NA

Weight Of Evidence Encoding

What?

In binary classification, replace categorical variables with

$$WOE_{ja} = \ln \frac{P(X_j = a | Y = 1)}{P(X_j = a | Y = 0)}$$

Why?

Leverage rich history in information theory and Bayesian statistics to manage overfitting of high cardinality variables

Cross Validation Target Encoding

Fold	Pay 1	Default Payment	Mean Target Encoding
2	Up To Date	0	0
3	Up To Date	0	0
2	Missed 1 Mo	1	0.5
3	Missed 1 Mo	0	0.5
Fold	Pay 1	Default Payment	CV Target Encoding
1	Up To Date	0	0
1	Missed 1 Mo	0	0.5
1	Missed 5 Mo	1	NA