



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

Weight Of Evidence Ending

what?

In binary classification, replace categorical variables with

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

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

why?

Weight Of Evidence Encoding

Pay 1	Default Payment	% 0s	% 1s	(% 1s) / (% 0s)	WOE
Up To Date	0	60 %	0 %	0	- Inf
Up To Date	0	60 %	0 %	0	- Inf
Up To Date	0	60 %	0 %	0	- Inf
Missed 1 Mo	1	40 %	50 %	1.25	0.223
Missed 1 Mo	0	40 %	50 %	1.25	0.223
Missed 1 Mo	0	40 %	50 %	1.25	0.223
Missed 5 Mo	1	0 %	50 %	Inf	Inf

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