



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

Weight Of Evidence Ending

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

Redundant Data

Real Data

Too much of the same is not a good thing

Solutions

- Features can be highly interrelated.
 - Not all features are related to the target.
-
- Leads to numeric instability in machine learning algorithms (GLM)
 - Overweights importance of redundant features (RF, GBM)
-
- Select representatives from clustered / grouped features
 - Use dimensionality reduction techniques (PCA, GLRM, MCA, MDS)

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