

DATA MINING CUP 2020

Features

Column name	Description	Value Range / Format
itemID	Unique identifier for every product	Positive integer
brand	Categorical attribute, different brands are deciphered by integer values	[0,...,274] / Positive integer (0 = no brand)
manufacturer	Categorical attribute, different characteristics are denoted by integers	[1,...,253] / Positive integer
customerRating	Numerical attribute, customer rating	[1,5] / Positive float (0.0 = no customer rating)
category1	Categorical and hierarchical attribute	[1,...,8] / Positive integer
category2	Categorical and hierarchical attribute	[1,...,52] / Positive integer
category3	Categorical and hierarchical attribute	[1,...,8] / Positive integer
recommendedRetailPrice	Numerical attribute, recommended retail price for a product, mostly provided by the manufacturer	Positive decimal number with two positions after the decimal point
salesPrice	Numerical attribute, actual price point of a product	Positive decimal number with two positions after the decimal point
order	Numerical attribute, number of sold units of a product in one transaction	Positive integer
transactID	Categorical attribute, ID to connect orders of different products by a single customer. The ID is not connected to a specific	Positive integer

	costumer in a fixed way. (i.e. a returning costumer gets every time a different ID). After an period of time an ID can reappear.	
time	Ordinal attribute, exact time stamp of a product purchase	YYYY-MM-dd hh:mm:ss
simulationPrice	Numerical attribute, the price of a product in the simulation period	Positive decimal number with two positions after the decimal point
promotion	“,” separated list of dates for products in promotion	Comma-separated list of dates (YYYY-MM-dd)