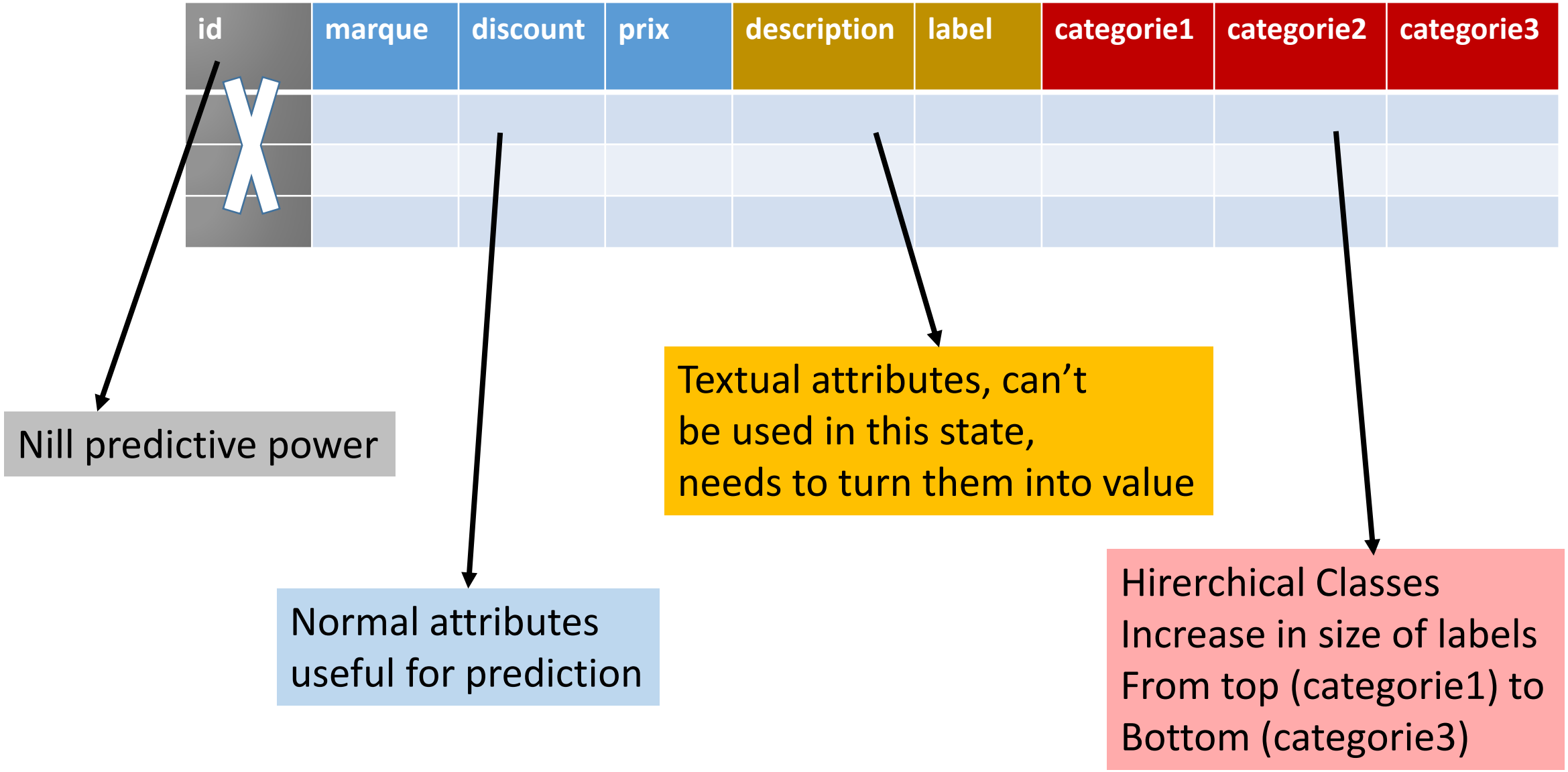
Hierarchical Classification (Internet Memory)

- A look into the nature of train and test data

Presented: M. Naeem

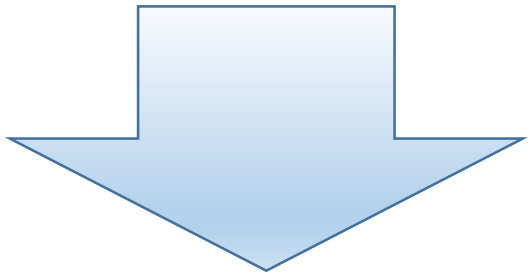

id
marque
discount
prix
description
label
categorie1
categorie2
categorie3
Nill predictive power
Normal attributes
useful for prediction
Textual attributes, can’t
be used in this state,
needs to turn them into value
Hirerchical Classes
Increase in size of labels
From top (categorie1) to
Bottom (categorie3)

Feature Engineering

- Generate some new attributes out of existing attributes in order to increase the predictabillity

description

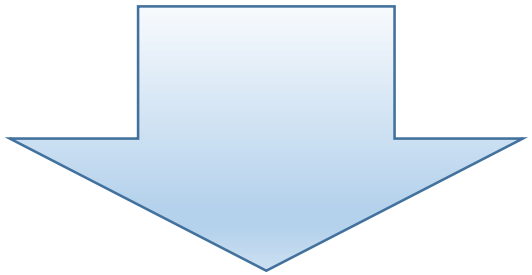
NLP Transormation
Input: text attribute
Output: variable number of numerical attributes



Description.1	Description.2	Description.3	Description.4	Description.5
0.00004564	0.0004784	0.0083694	0.00144564	0.000012364
.	.	.	.	.
.	.	.	.	.

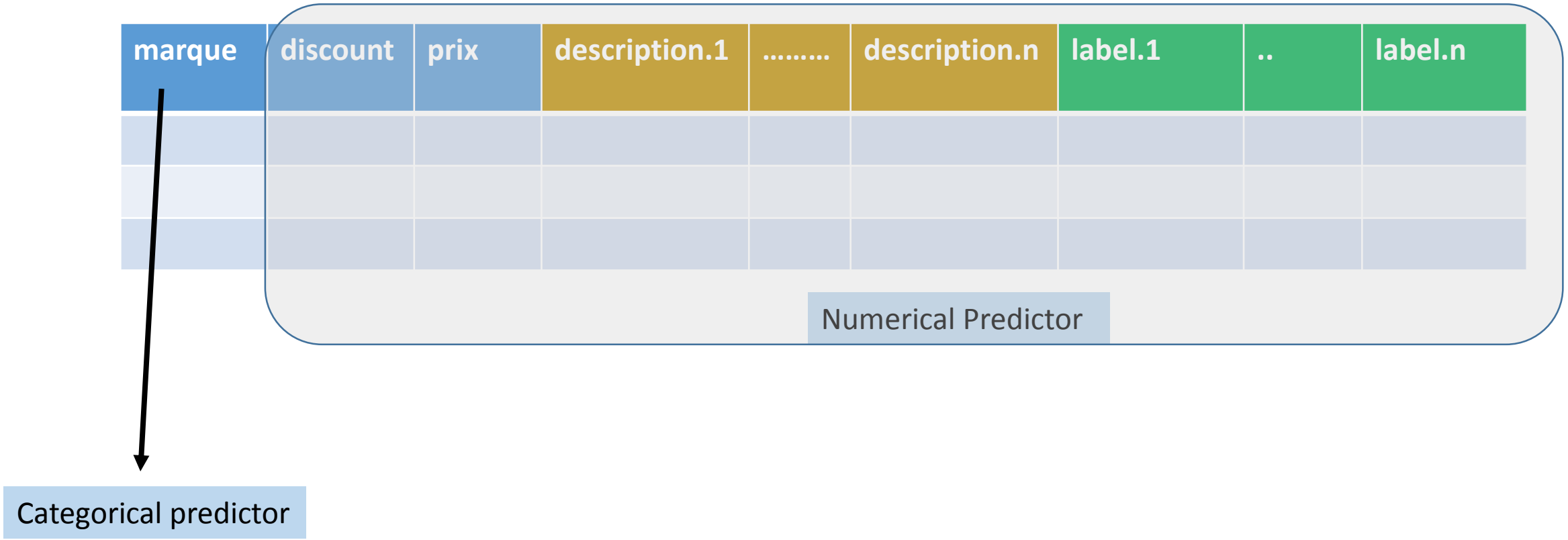
label

NLP Transormation
Input: text attribute
Output: variable number of numerical attributes



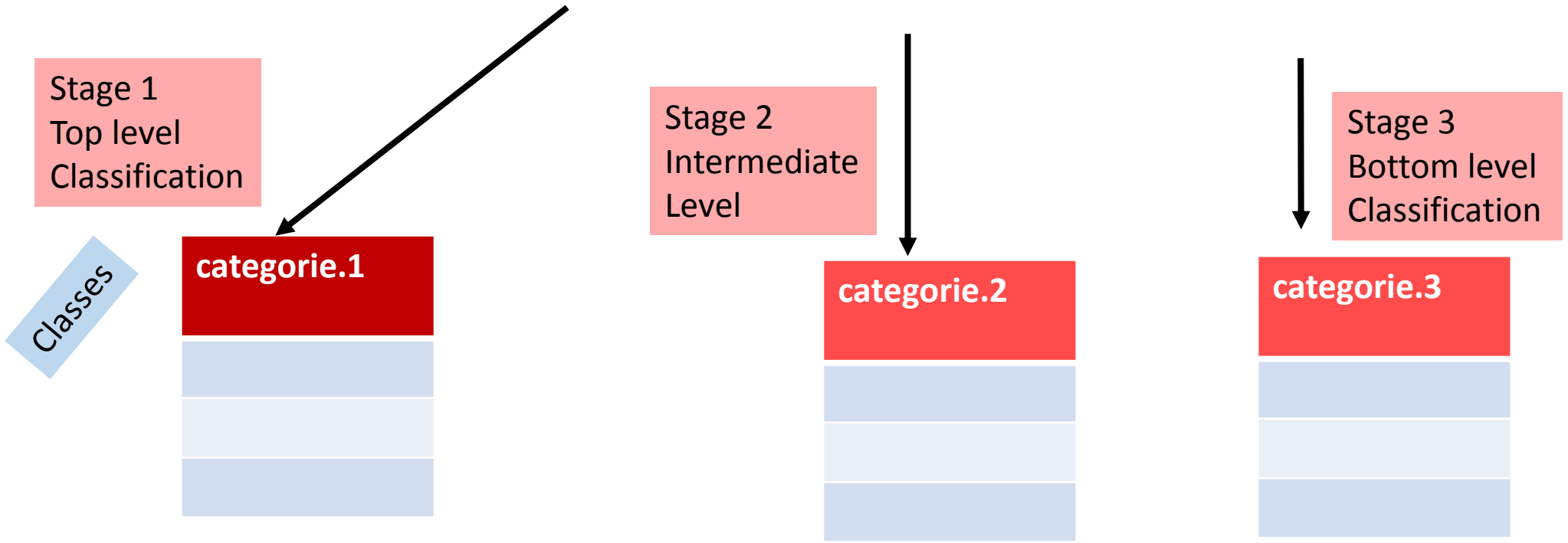
Description.1	Description.2	Description.3	Description.4	Description.5
0.00004564	0.0004784	0.0083694	0.00144564	0.000012364
.	.	.	.	.
.	.	.	.	.

Train Data after data engineering



marque	discount	prix	description.1	……	description.n	label.1	..	label.n

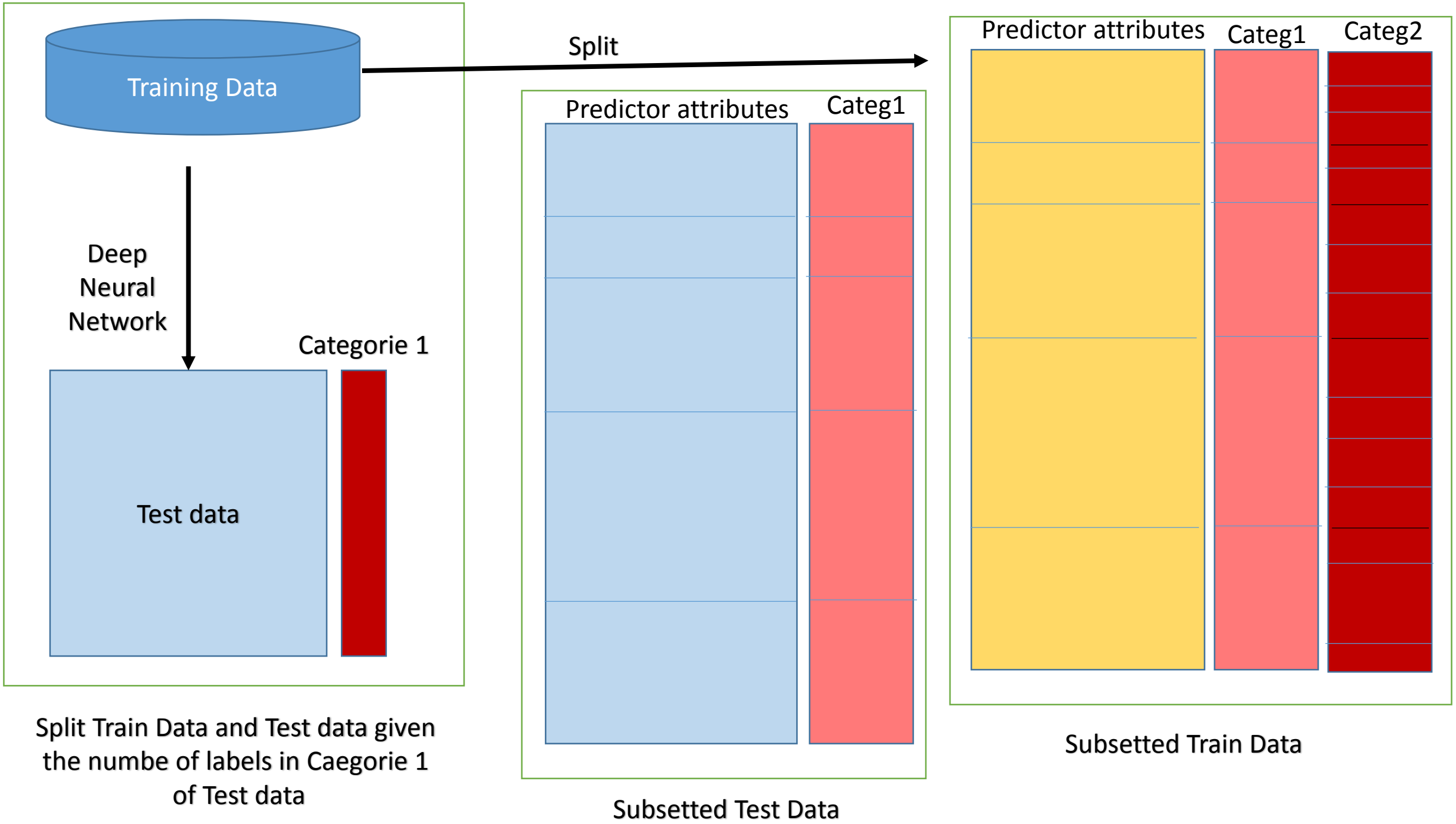
Predictor



Stage 1 (Prediction of Categorie 1)

- Classical approach of learning data and then predicting the labels of categorie 1

- Subset the train data and test data given the categories in the outcome predicted labels

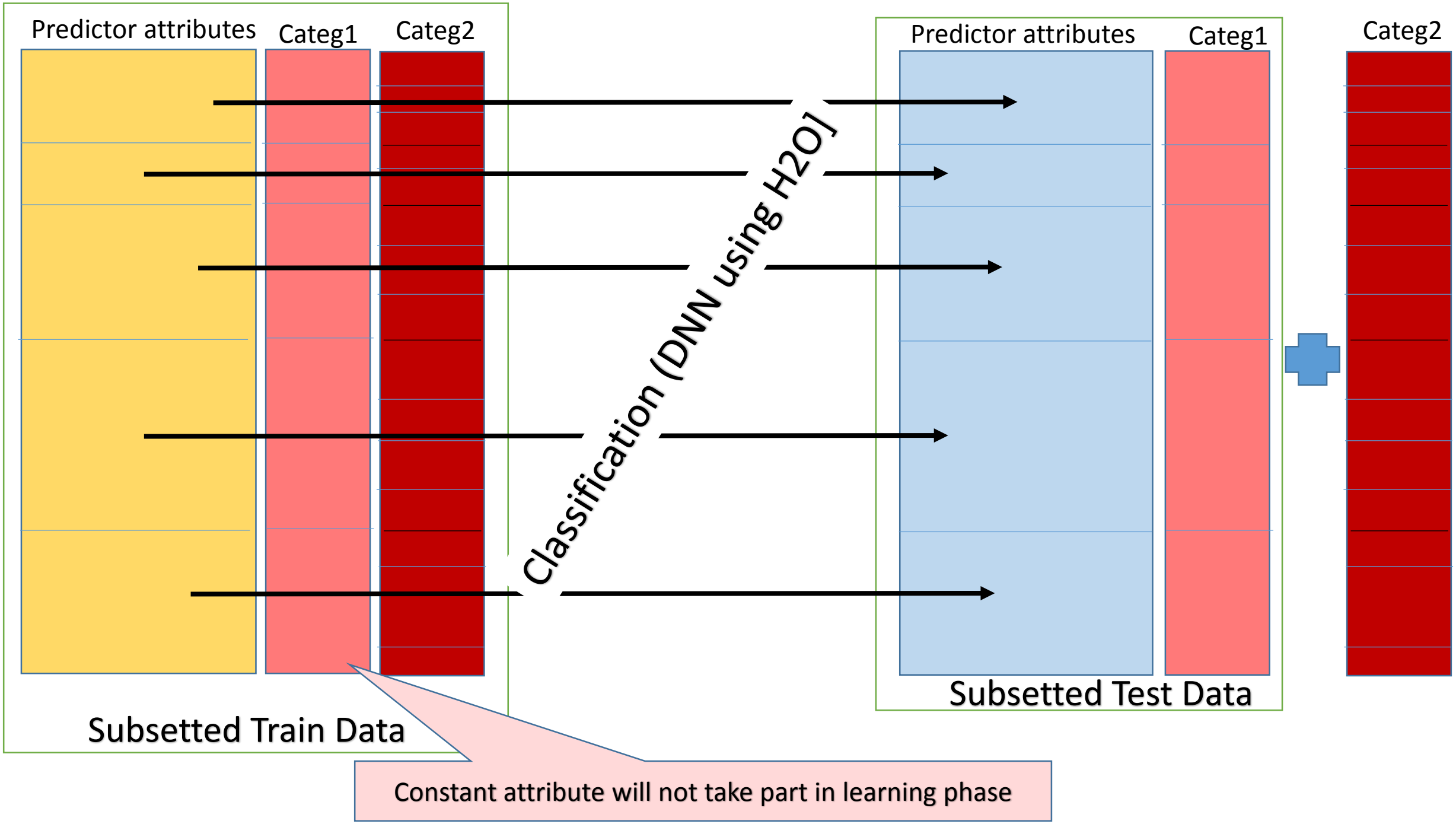

Training Data
Split
Predictor attributes
Categ1
Categ2
Deep
Neural
Network
Categorie 1
Test data
Predictor attributes
Categ1
Split Train Data and Test data given the numbe of labels in Caegorie 1 of Test data
Subsetted Test Data
Subsetted Train Data

Stage 2 (Prediction of Categorie 2)

- Learn each of the subsetted train data and then predict the labels of corresponding test data

- At the end of round of learning/prediction of all sets of train data, all predicted outcome are ensembled in one attribute

- Again subset the train data and test data given the categories 1 and 2 in the outcome predicted labels


Predictor attributes
Categ1
Categ2
Classification (DNN using H2O)
Predictor attributes
Categ1
Categ2
Subsetted Test Data
Subsetted Train Data
Constant attribute will not take part in learning phase

Stage 3 (Prediction of Categorie 3)

- We shall subset the train data as per outcomeof categorie1 and catetorie2 of test data

- Repeat the same process doing ‘x’ number of learning in DNN classifier

End