

Topics in Business Intelligence

Case study comparison and feedback

Tommi Tervonen

Econometric Institute, Erasmus University Rotterdam

Feedback (+)

Group	+
k-NN and NB	Discretization in NB, Clear explanation of k-NN
ctree	Very good method explanation No strange “moving parts”
logreg	Good method explanation, Nice use of PCA to find the two groups of attributes
NN	Forward propagation clear, No overfitting in analysis

Feedback (-)

Group	-
k-NN and NB	Lacking some math, very strange discretization (why for k-NN?)
ctree	No asymmetric misclassification costs, 2 models
logreg	Used strange stuff from SPSS without explaining the methods, outlier analysis
NN	Learning algorithm not explained properly, weak biological interpretation / justification

Method	-
k-NN	ANNDON + TIMELR, k=20
NB	Discretization by frequency, ANNDON + FRQRES
ctree	Stratified oversampling 7:1, all attr.
logreg	TIMELR, FRQRES, LSTDON, ANNDON
NNet	PCA, 4 PCs with largest eigenvalues (90% of trace)

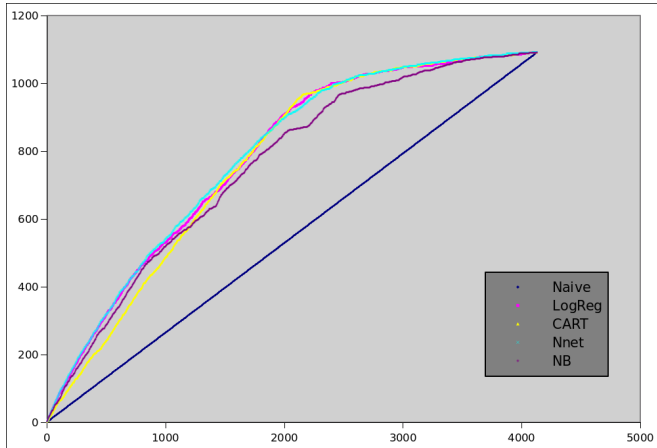
Sensitivity and specificity

Group	Sensitivity	Specificity
k-NN	0%	0%
NB	89%	51%
ctree	92%	50%
logreg	92%	51%
NNet	95%	39%

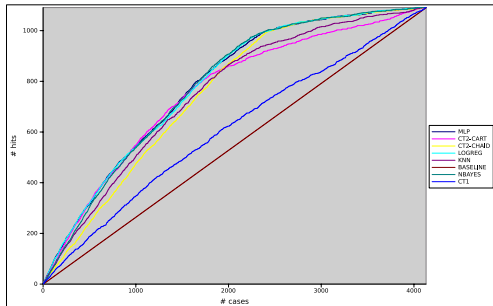
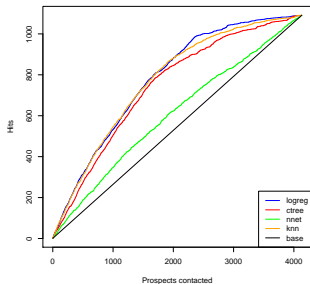
Comparison of models

- Data test and training data used is from year 1998
- Let's validate the models with data from 1999!

Comparison of models



Comparison of models (previous years)



UNDERSTANDING ONLINE STAR RATINGS:

★★★★★ [HAS ONLY ONE REVIEW]

★★★★★ EXCELLENT

★★★★☆ OK

★★★★☆

★★★★☆

★★★☆☆

★★★☆☆

★★☆☆☆

★☆☆☆☆



CRAP