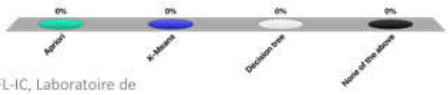
Which data mining algorithm belongs to the Expectation-Maximization class?

A. Apriori

B. K-Means

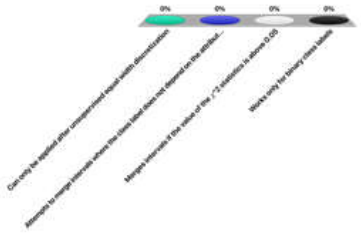
C. Decision tree

D. None of the above



Supervised discretization …

A. Can only be applied after unsupervised equal width discretization
B. Attempts to merge intervals where the class label does not depend on the attribute value distribution
C. Merges intervals if the value of the χ^2 statistics is above 0.05
D. Works only for binary class labels



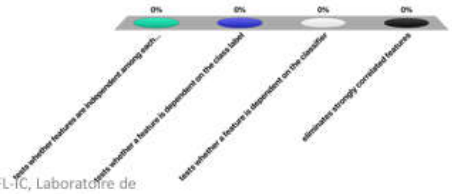
The filtering approach to feature selection …

A. tests whether features are independent among each other

B. tests whether a feature is dependent on the class label

C. tests whether a feature is dependent on the classifier

D. eliminates strongly correlated features



Which is the "best" classifier?

Classifier 1

Classified	Class: Cancer	Class: ¬Cancer
Cancer	45	20
¬Cancer	5	30

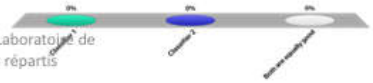
Classifier 2

Classified	Class: Cancer	Class: ¬Cancer
Cancer	40	10
¬Cancer	10	40

A. Classifier 1

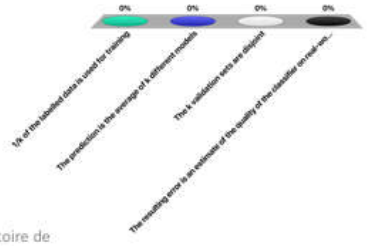
B. Classifier 2

C. Both are equally good



In k-fold cross-validation …

A. 1/k of the labelled data is used for training
B. The prediction is the average of k different models
C. The k validation sets are disjoint
D. The resulting error is an estimate of the quality of the classifier on real-world data



Which is wrong?

A. The lower model complexity, the higher bias
B. The higher model complexity, the higher variance
C. The higher the data volume, the higher the training error
D. The training error is always higher than the test error