User manual

Made by : ROBLIN Caroline, TELLIER Kevin, YE Maxime

06 novembre 2019

Data presentation

Comparison of SVM and another model

Inputs

Model to compare with svm

- Logistic Regression
- Decision Tree
- Random Forest
- Gradient Boosting
- XGBoost

- for the Decision Tree :
  Minsplit : represents the minimum number of observations in a node for a split to take place (5)
  Minbucket : says the minimum number of observations I should keep in terminal nodes (15)
  Cp : it's the complexity parameter (0.001)
- for the Random Forest :
  Number of trees (157)
  Node Size (12)
  Mtry (9)
- for the Gradient Boosting :
  N trees (256)
  interaction depth (5)
  Min obs in node : refers to the minimum number of observations in a tree node (33)
  shrinkage : it's the regulation parameter which dictates how fast / slow the algorithm should move (0.24).
- for the XGBoost :
  Nround (256)
  Max depth (20)
  Lambda (0.56)

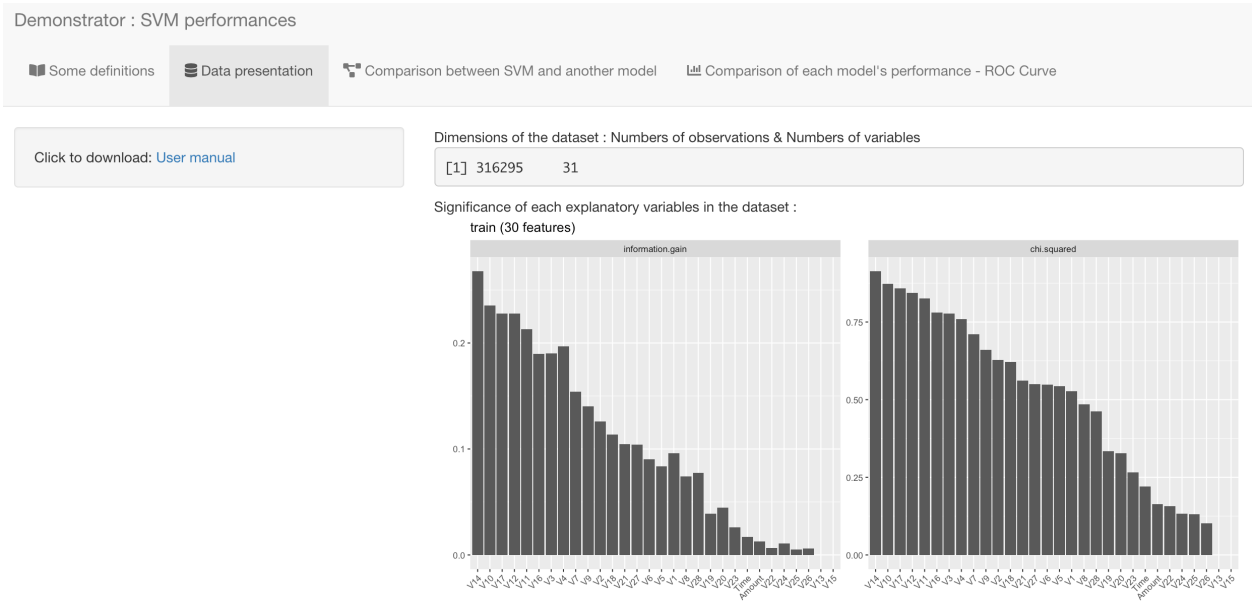

Figure 1: Data presentation

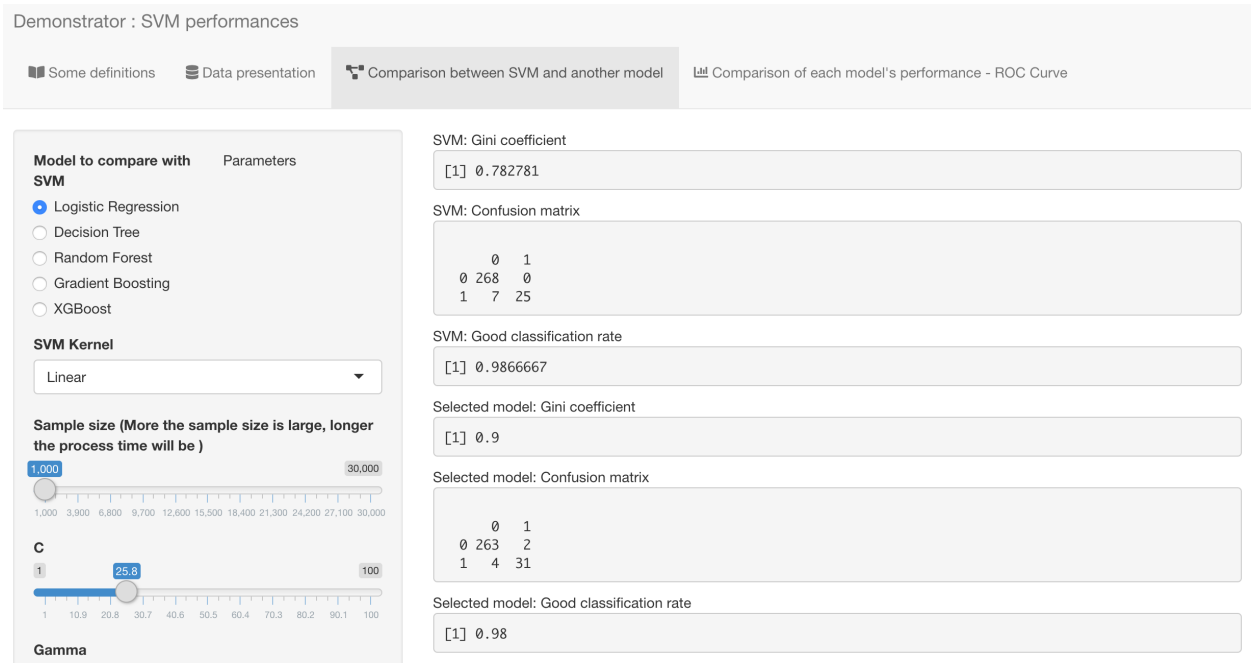

Figure 2: Comparing of SVM with another Machine Learning model

Eta (0.278)
Sub sample (0.56)
Min child weight (4)
Cold sample by tree (0.683)

SVM Kernel

- Linear
- Polynomial
- Radial Basis
- Sigmoid

Sample size (the larger the size chosen, the longer the processing time will be)

C

Gamma

Outputs

- SVM : Gini coefficient
- SVM : confusion matrix
- SVM : good classification rate
- Selected model : Gini coefficient
- Selected model : Confusion matrix
- Selected model : good classification rate
- ROC Curve comparison between the SVM and the selected model

Comparison of each model's performance - ROC Curve

Inputs

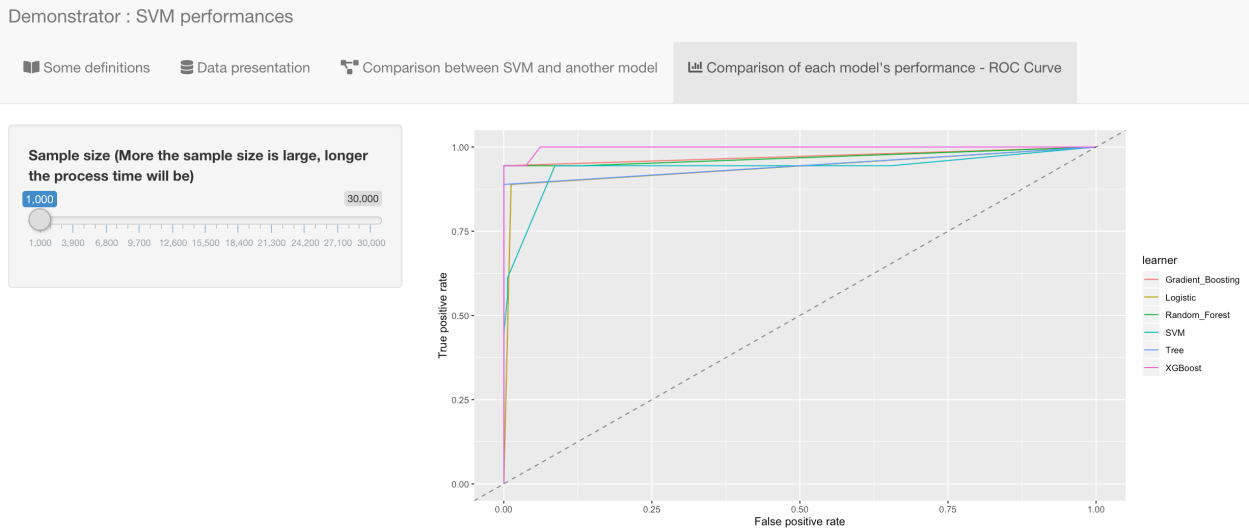

Figure 3: Comparing model performance

Sample size (the larger the size chosen,the longer the processing time will be)

Outputs

You can see that it isn't easy to choose the best model regardless of the sample size because of the crossing. It's better to refer to the Gini index or the good classification rate we have in the previous tab.