



UNIWIN VERSION 10.6.0

RENFORCEMENT DU GRADIENT

Révision : 07/07/2026

Définition.....	1
Entrée des données	2
Données manquantes	3
Exemple 1 : Classement binaire - Fichier Infarct2	4
L'option Rapports	7
L'option Graphiques	10
Exemple 2 : Classement multi-classes - Fichier Iris3	15
Exemple 3 : Régression - Fichier Boston	22
Enregistrement des résultats	28
Références	29

Définition

La procédure Renforcement du gradient (gradient boosting) implémente une méthode d'apprentissage machine pour prévoir des observations à partir de données. Elle crée des modèles de deux formes :

1. Modèles de classement qui divisent des observations en classes en se basant sur les caractéristiques observées.
2. Modèles de régression qui prévoient la valeur d'une variable cible.

Elle consiste à combiner progressivement des modèles faibles basés sur des arbres de décision ou de régression pour former un modèle prédictif performant en corrigeant les erreurs résiduelles de manière itérative. Cette méthode repose sur une optimisation séquentielle où chaque nouveau modèle ajuste ses paramètres en minimisant une fonction de perte via le calcul des gradients.

Les observations sont classiquement divisées en trois jeux : un jeu d'apprentissage utilisé pour construire le modèle, un jeu de validation, pour lequel la classe ou la valeur est connue, utilisé pour valider le modèle et un jeu de prévision, pour lequel la classe ou la valeur n'est pas connue, utilisé pour faire les prévisions désirées.

La variable cible et les caractéristiques prédictives peuvent être qualitatives ou quantitatives.

Les données brutes sont utilisées pour les calculs car la structure des arbres de décision et de régression n'est pas impactée par les habituelles transformations monotones des données.

Cette procédure est basée sur le package R 'gbm' (Gradient Boosting Machine).

Entrée des données

Cliquons sur l'icône GBOOST dans le ruban Expliquer pour afficher la boîte de dialogue montrée ci-dessous.

The screenshot shows a dialog box titled "Renforcement du gradient" with a close button (X) in the top right corner. The dialog is divided into several sections:

- Variable à expliquer :** A text input field with a right-pointing arrow icon to its left.
- Variables explicatives quantitatives :** A list box with an up/down arrow icon to its left.
- Variables explicatives qualitatives :** A list box with an up/down arrow icon to its left.
- Méthode :** Two radio buttons: "Classement" (selected) and "Régression".
- Nombre maximum d'arbres :** A text input field containing "100".
- Profondeur maximale des arbres :** A text input field containing "3".
- Taille minimum des noeuds terminaux :** A text input field containing "10".
- Proportion des observations :** A text input field containing "0,5".
- Taux d'apprentissage :** A text input field containing "0,1".
- Nombre de validations croisées :** A text input field containing "5".
- Racine aléatoire :** A text input field containing "16807".
- (Poids des observations :) :** A text input field with an up/down arrow icon to its left.
- (Libellés des variables quantitatives :) :** A text input field with an up/down arrow icon to its left.
- (Libellés modalités des variables qualitatives :) :** A text input field with an up/down arrow icon to its left.
- (Libellés des observations :) :** A text input field with an up/down arrow icon to its left.

At the bottom of the dialog, there are five buttons: "Ok", "Annuler", "Sélection", "Supprimer", and "Aide".

Cette boîte de dialogue permet de préciser la variable à expliquer ainsi que les variables explicatives quantitatives (s'il y en a) et qualitatives (s'il y en a) à utiliser.

Elle permet également, en option, d'indiquer le nom de la variable contenant les poids des observations, le nom de la variable contenant les libellés des variables quantitatives, le nom de la variable contenant les libellés des modalités des variables qualitatives ainsi que le nom de la variable contenant les libellés des observations.

Le choix de la méthode est proposé : Classement (si la variable à expliquer est qualitative) ou Régression (si la variable à expliquer est quantitative).

La distribution utilisée pour la variable à expliquer est automatiquement sélectionnée : Bernoulli si classement binaire, multinomiale si classement à plus de 2 classes et gaussienne si régression.

Les autres paramètres sont :

- Nombre maximum d'arbres à ajuster (nombre d'itérations)
- Profondeur maximale des arbres
- Taille minimum des nœuds terminaux
- Proportion des observations d'apprentissage utilisée pour bâtir les arbres (sélection aléatoire des données du jeu d'apprentissage à chaque itération, c'est le « stochastic gradient boosting » de Friedman)
- Taux d'apprentissage
- Nombre de validations croisées pour déterminer le nombre optimal d'arbres
- Racine aléatoire

En résumé :

Pour chaque validation croisée :

- └─ Pour chaque itération $m = 1, \dots$, nombre maximum d'arbres :
 - └─ Tirer aléatoirement proportion des observations $\times$ nombre d'observations
 - └─ Calculer les pseudo-résidus
 - └─ Ajuster un arbre sur le sous-échantillon
- └─ Mettre à jour le modèle

Données manquantes

Les observations ayant des données manquantes pour les variables explicatives sont automatiquement éliminées des calculs.

Les observations ayant des données manquantes pour la variable à expliquer constituent le jeu de prévision.

Exemple 1 : Classement binaire - Fichier Infarct2

Ce fichier contient des informations concernant 101 victimes d'un infarctus du myocarde.
Les variables mesurées sont :

frcar fréquence cardiaque

insys index systolique

papul pression artérielle pulmonaire

repul résistance pulmonaire

incar index cardiaque

prdia pression diastolique

pvent pression ventriculaire

La variable *libobs* contient les libellés des observations et la variable qualitative *groupe* indique par ses deux codes les personnes décédées ou vivantes.

Sélectionnons la variable *groupe* comme variable à expliquer, les variables *frcar* à *repul* comme variables explicatives quantitatives, la variable *libobs* pour les libellés des observations, choisissons la méthode *Classement* et laissons les autres paramètres aux valeurs par défaut.

Renforcement du gradient

Variable à expliquer :
groupe

Variables explicatives quantitatives :
frcar
incar
insys
prdia
papul
pvent
repul

Variables explicatives qualitatives :

Méthode :
☒ Classement ☐ Régression

Nombre maximum d'arbres : 100

Profondeur maximale des arbres : 3

Taille minimum des noeuds terminaux : 10

Proportion des observations : 0,5

Taux d'apprentissage : 0,1

Nombre de validations croisées : 5

Racine aléatoire : 16807

(Poids des observations :)

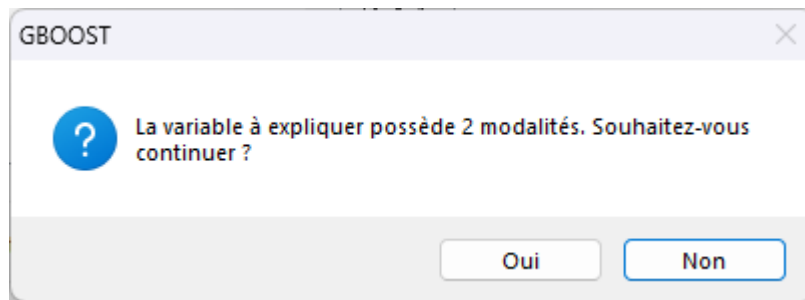
(Libellés des variables quantitatives :)

(Libellés modalités des variables qualitatives :)

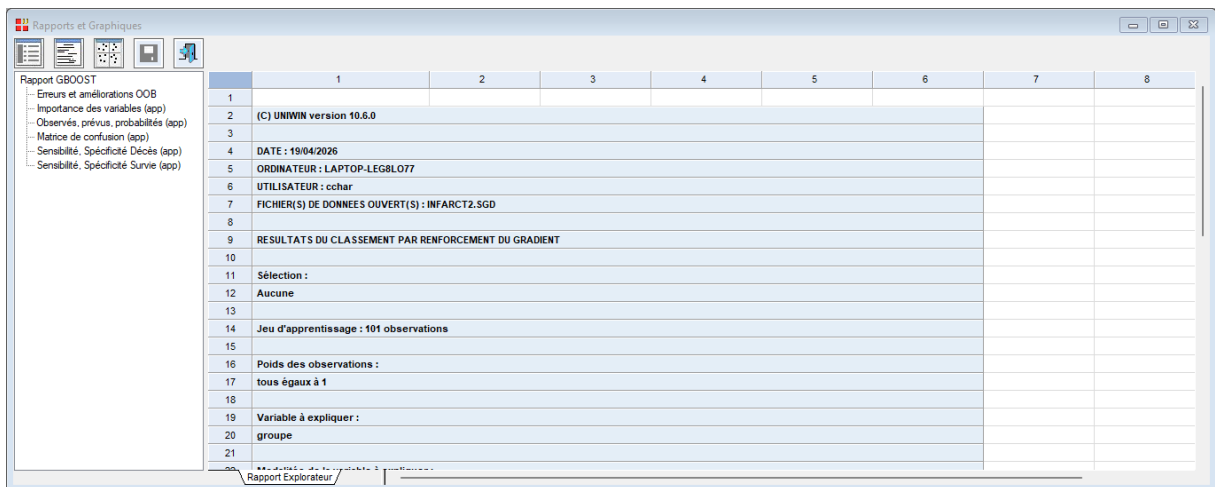
(Libellés des observations :)
libobs

Ok Annuler Sélection Supprimer Aide

Cliquons sur OK et confirmons que la méthode de classement utilise une variable à 2 modalités.

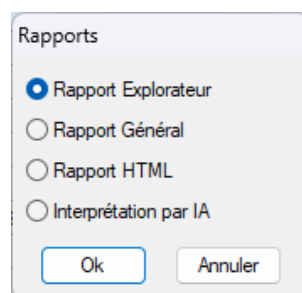


Après quelques instants, la fenêtre suivante s'affiche :



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



et l'icône 'Graphiques'  affiche la boîte de dialogue des options pour les graphiques.

Graphiques

- ☒ Graphique des erreurs (apprentissage, validation croisée)
- ☐ Graphique de l'importance des variables explicatives
- ☐ Nuage de points (apprentissage)
- ☐ Nuage de points (validation)
- ☐ Nuage de points (prévision)
- ☐ Graphique des frontières (apprentissage)
- ☐ Graphique des frontières (validation)
- ☐ Diagramme de la matrice de confusion (apprentissage)
- ☐ Diagramme de la matrice de confusion (validation)
- ☐ Courbes ROC (apprentissage)
- ☐ Courbes ROC (validation)

Ok Annuler

L'icône 'Enregistrer'  permet de sélectionner les résultats de l'analyse à enregistrer dans un fichier.

Enregistrement des résultats (1/2)

Enregistrer

- ☐ Erreurs d'apprentissage
- ☐ Erreurs de validation croisée
- ☐ Améliorations OOB
- ☐ Variables explicatives
- ☐ Importance des variables
- ☐ Libellés des observations d'apprentissage
- ☐ Probabilités pour le jeu d'apprentissage
- ☐ Valeurs observées pour le jeu d'apprentissage
- ☐ Valeurs prévues pour le jeu d'apprentissage
- ☐ Seuils (apprentissage)

Noms attribués aux variables cibles

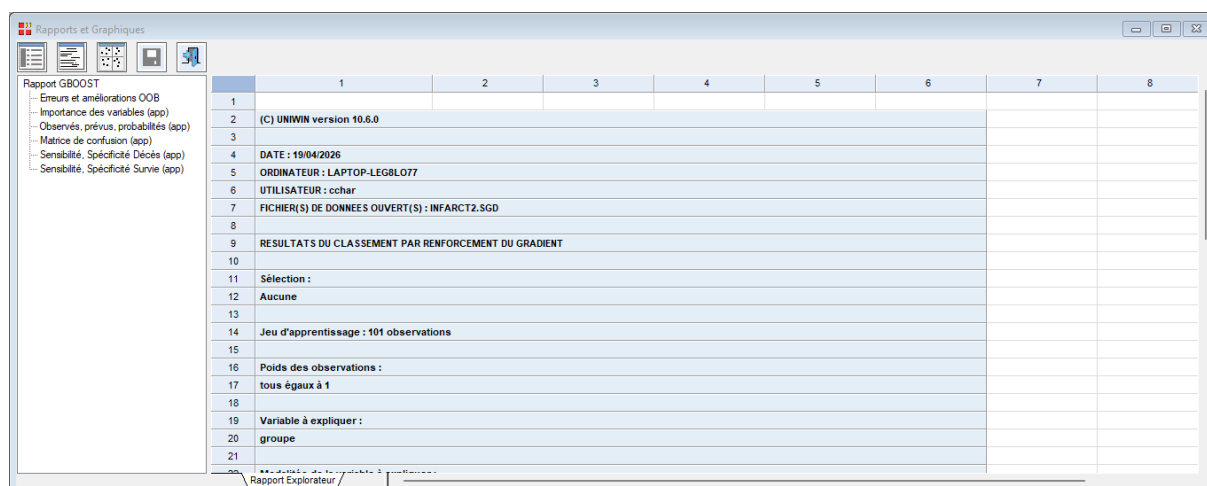
erreurapp
erreurcv
amelooob
varexpli
impvarexpli
libobsapp
probapp_1
obsapp
vpapp
seuilA

Ok Plus Tout Annuler

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme Explorateur, Général ou HTML et d'accéder à l'aide à l'interprétation par Intelligence Artificielle.

Le premier tableau indique que le jeu d'apprentissage est composé de 101 observations et rappelle les modalités de la variable à expliquer, les variables explicatives utilisées et les paramètres de l'analyse dont la loi utilisée pour la variable à expliquer, ici Bernoulli.

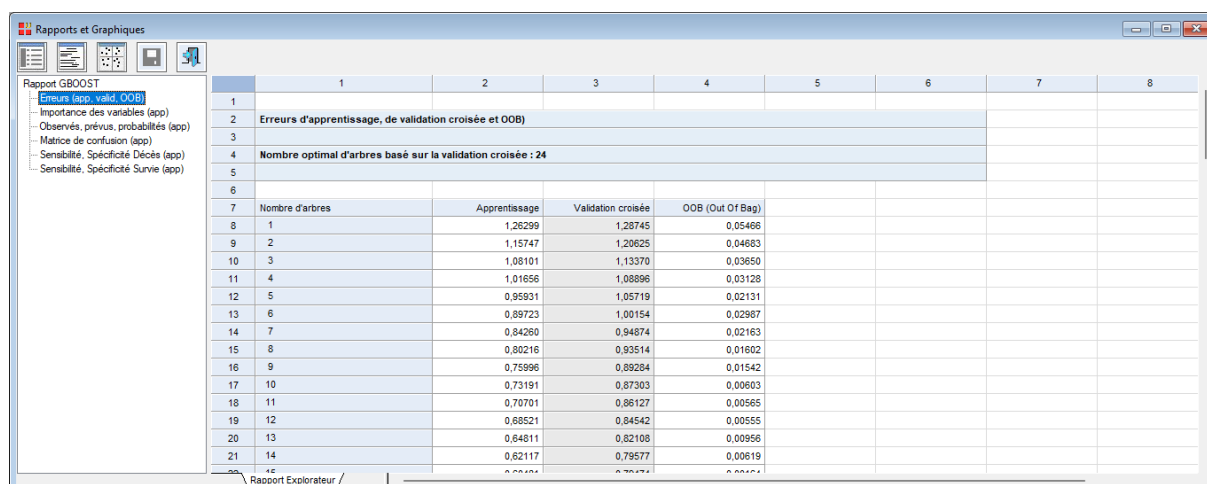


	1	2	3	4	5	6	7	8
1								
2	(C) UNIWIN version 10.6.0							
3								
4	DATE : 19/04/2026							
5	ORDINATEUR : LAPTOP-LEG8LO77							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNEES OUVERT(S) : INFARCT2.SGD							
8								
9	RESULTATS DU CLASSEMENT PAR RENFORCEMENT DU GRADIENT							
10								
11	Sélection :							
12	Aucune							
13								
14	Jeu d'apprentissage : 101 observations							
15								
16	Poids des observations :							
17	tous égaux à 1							
18								
19	Variable à expliquer :							
20	groupe							
21								

Le deuxième tableau indique pour chaque itération (nombre d'arbres ajustés) l'évolution de la fonction de perte (log-vraisemblance) évaluée en utilisant soit les données du jeu d'apprentissage, soit par validation croisée, soit avec la proportion des données non utilisées (OOB = Out Of Bag) du jeu d'apprentissage.

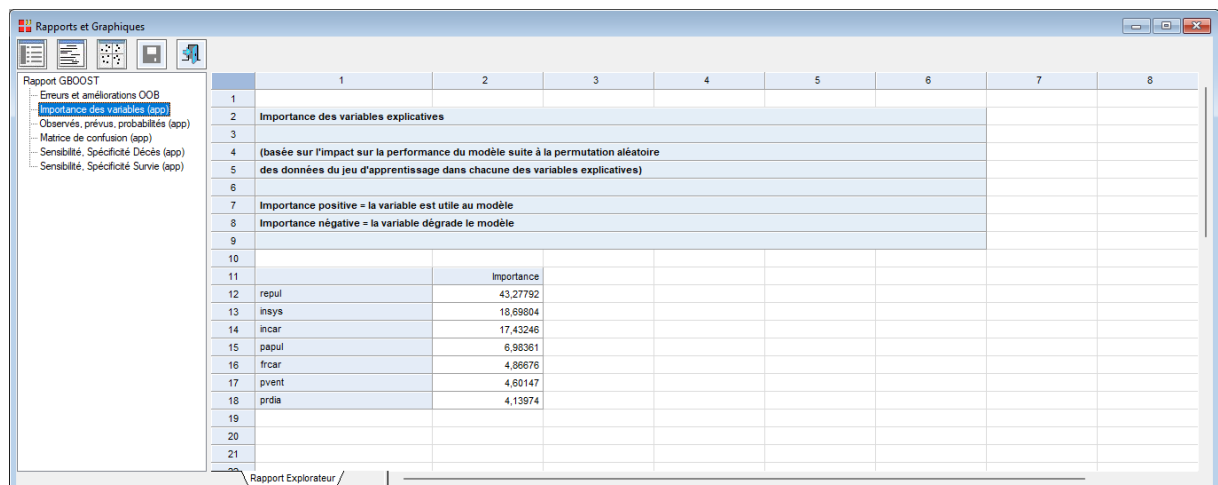
Il précise également le nombre optimal d'arbres en se basant sur la validation croisée. Il est ici de 24.

Ce nombre optimal sera utilisé pour la suite des calculs.



	1	2	3	4	5	6	7	8
1								
2	Erreurs d'apprentissage, de validation croisée et OOB							
3								
4	Nombre optimal d'arbres basé sur la validation croisée : 24							
5								
6								
7	Nombre d'arbres	Apprentissage	Validation croisée	OOB (Out Of Bag)				
8	1	1,26299	1,28745	0,05466				
9	2	1,15747	1,20625	0,04683				
10	3	1,08101	1,13370	0,03650				
11	4	1,01656	1,08896	0,03128				
12	5	0,95931	1,05719	0,02131				
13	6	0,89723	1,00154	0,02967				
14	7	0,84280	0,94874	0,02163				
15	8	0,80216	0,93514	0,01602				
16	9	0,75996	0,89284	0,01542				
17	10	0,73191	0,87303	0,00603				
18	11	0,70701	0,86127	0,00565				
19	12	0,68521	0,84542	0,00555				
20	13	0,64811	0,82108	0,00956				
21	14	0,62117	0,79577	0,00619				

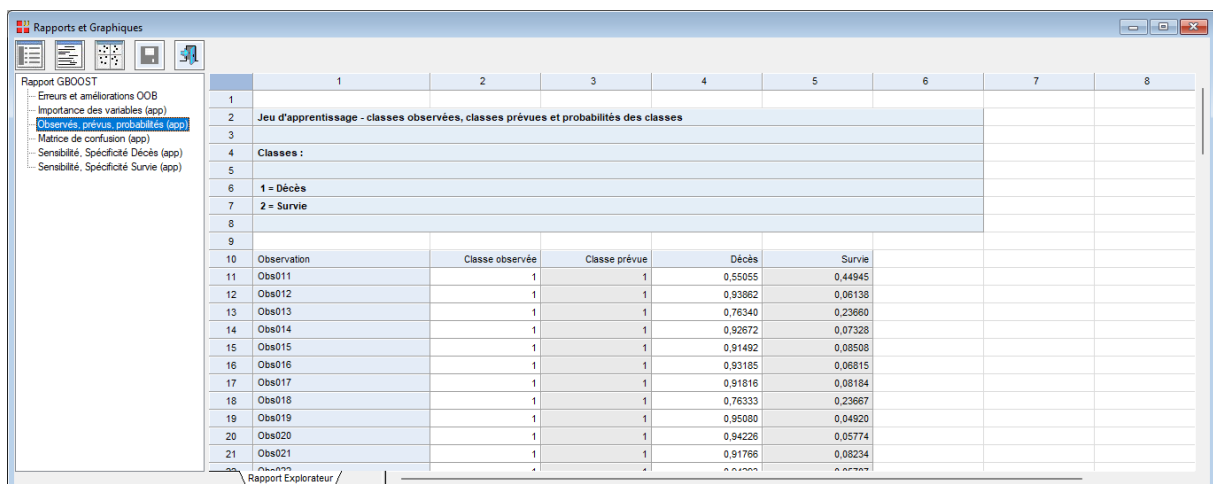
Le troisième tableau affiche l'importance des variables explicatives dans le modèle.



	1	2	3	4	5	6	7	8
1								
2	Importance des variables explicatives							
3								
4	(basée sur l'impact sur la performance du modèle suite à la permutation aléatoire des données du jeu d'apprentissage dans chacune des variables explicatives)							
5								
6								
7	Importance positive = la variable est utile au modèle							
8	Importance négative = la variable dégrade le modèle							
9								
10								
11		Importance						
12	regul	43,27792						
13	insys	18,69804						
14	incar	17,43246						
15	papul	6,96361						
16	frcar	4,86676						
17	pvent	4,60147						
18	prdia	4,13974						
19								
20								
21								

L'importance est basée sur l'impact sur la performance du modèle à la suite de la permutation aléatoire des données du jeu d'apprentissage dans chacune des variables explicatives.

Le quatrième tableau affiche pour chaque observation du jeu d'apprentissage les valeurs observées et prévues de la variable à expliquer ainsi que les probabilités d'affectation à chacune des classes.



	1	2	3	4	5	6	7	8
1								
2	Jeu d'apprentissage - classes observées, classes prévues et probabilités des classes							
3								
4	Classes :							
5								
6	1 = Décès							
7	2 = Survie							
8								
9								
10	Observation	Classe observée	Classe prévue	Décès	Survie			
11	Obs011	1	1	0,55055	0,44945			
12	Obs012	1	1	0,93862	0,06138			
13	Obs013	1	1	0,76340	0,23660			
14	Obs014	1	1	0,92672	0,07328			
15	Obs015	1	1	0,91492	0,08508			
16	Obs016	1	1	0,93185	0,06815			
17	Obs017	1	1	0,91816	0,08184			
18	Obs018	1	1	0,76333	0,23667			
19	Obs019	1	1	0,95080	0,04920			
20	Obs020	1	1	0,94226	0,05774			
21	Obs021	1	1	0,91766	0,08234			

Le tableau suivant affiche la matrice de confusion pour le jeu d'apprentissage.

Environ 92% des observations sont bien classées par le modèle construit.

Les deux tableaux suivants affichent les sensibilités et spécificités ainsi que l'aire sous la courbe ROC pour chacune des classes.

Ces tableaux ne sont disponibles que pour un classement binaire (deux modalités pour la variable à expliquer).

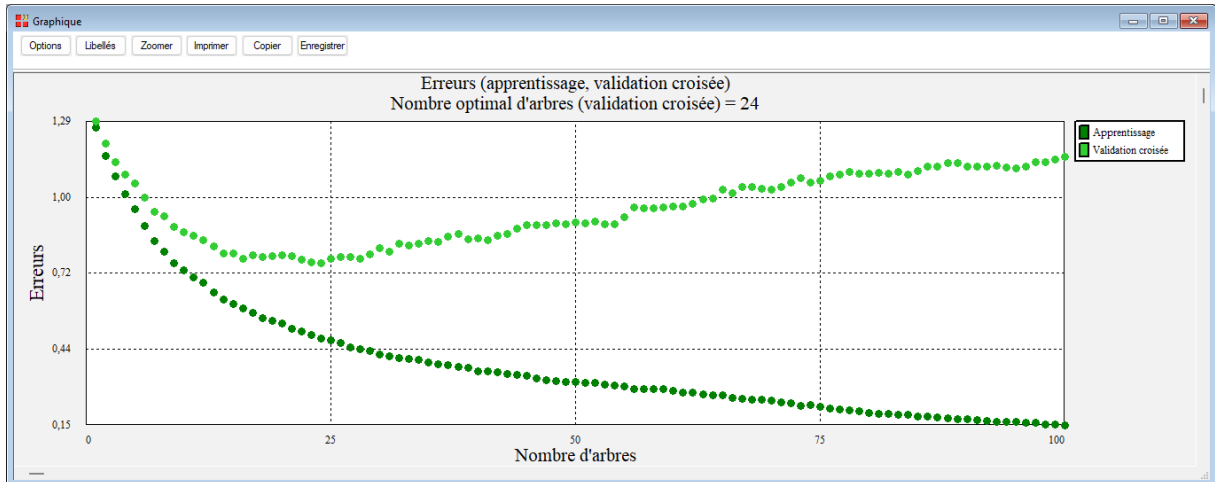
Rapports et Graphiques								
Rapport GBOOST								
Erreurs et améliorations OOB								
Importance des variables (app)								
Observés, prévus, probabilités (app)								
Matrice de confusion (app)								
Sensibilité, Spécificité Décès (app)								
Sensibilité, Spécificité Survie (app)								
1	1	2	3	4	5	6	7	8
2	Matrice de confusion pour le jeu d'apprentissage							
3								
4	Basée sur le tableau des prévisions							
5								
6	En lignes, les classes observées							
7	En colonnes, les classes prévues							
8								
9	Pourcentage de mal classés : 7,921 %							
10	Pourcentage de bien classés (exactitude) : 92,079 %							
11								
12	Précision = $VP / (VP + FP)$							
13	Rappel = $VP / (VP + FN)$							
14	Score F1 = $2 \times (\text{Précision} \times \text{Rappel}) / (\text{Précision} + \text{Rappel})$							
15								
16								
17	Observé \ Prévu	Taille	Décès	Survie	Précision	Rappel	Score F1	
18	Décès	51	47	4	0,92157	0,92157	0,92157	
19	Survie	50	4	46	0,92000	0,92000	0,92000	
20								
21								

Rapports et Graphiques								
Rapport GBOOST								
Erreurs et améliorations OOB								
Importance des variables (app)								
Observés, prévus, probabilités (app)								
Matrice de confusion (app)								
Sensibilité, Spécificité Décès (app)								
Sensibilité, Spécificité Survie (app)								
1	1	2	3	4	5	6	7	8
2	SENSIBILITE, SPECIFICITE POUR LE JEU D'APPRENTISSAGE							
3								
4	Classe : Décès							
5								
6	Sensibilité en %							
7	Spécificité en %							
8								
9	Aire sous la courbe (AUC) = 0,977							
10								
11								
12		Seuil	Sensibilité	Spécificité				
13	1	-Infini	100,00000	0				
14	2	0,05100	100,00000	0				
15	3	0,05670	100,00000	10				
16	4	0,06145	100,00000	12				
17	5	0,06283	100,00000	14				
18	6	0,06916	100,00000	16				
19	7	0,07259	100,00000	18				
20	8	0,07549	100,00000	20				
21	9	0,07703	100,00000	22				

Rapports et Graphiques								
Rapport GBOOST								
Erreurs et améliorations OOB								
Importance des variables (app)								
Observés, prévus, probabilités (app)								
Matrice de confusion (app)								
Sensibilité, Spécificité Décès (app)								
Sensibilité, Spécificité Survie (app)								
1	1	2	3	4	5	6	7	8
2	SENSIBILITE, SPECIFICITE POUR LE JEU D'APPRENTISSAGE							
3								
4	Classe : Survie							
5								
6	Sensibilité en %							
7	Spécificité en %							
8								
9	Aire sous la courbe (AUC) = 0,977							
10								
11								
12		Seuil	Sensibilité	Spécificité				
13	1	-Infini	0	100,00000				
14	2	0,05100	0	100,00000				
15	3	0,05670	10	100,00000				
16	4	0,06145	12	100,00000				
17	5	0,06283	14	100,00000				
18	6	0,06916	16	100,00000				
19	7	0,07259	18	100,00000				
20	8	0,07549	20	100,00000				
21	9	0,07703	22	100,00000				

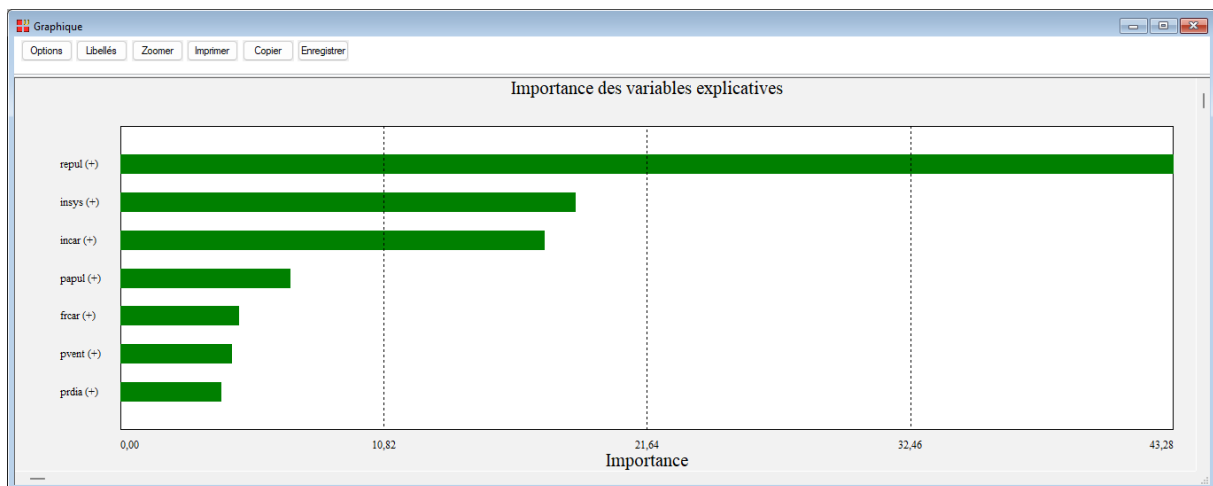
L'option Graphiques

- Graphique des erreurs (apprentissage et validation croisée)



Le nombre optimal d'arbres à ajuster déterminé par validation croisée est de 24.

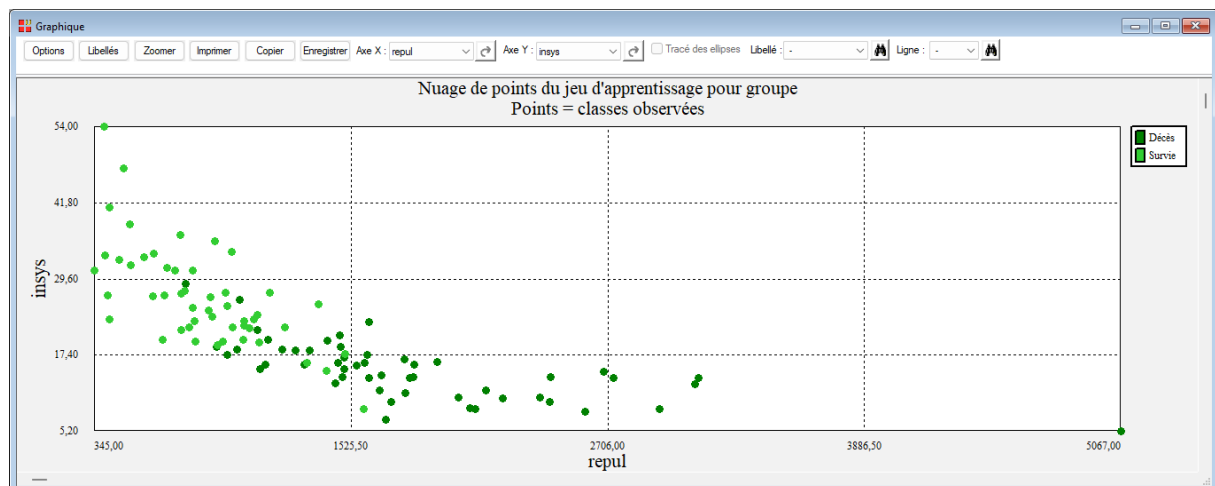
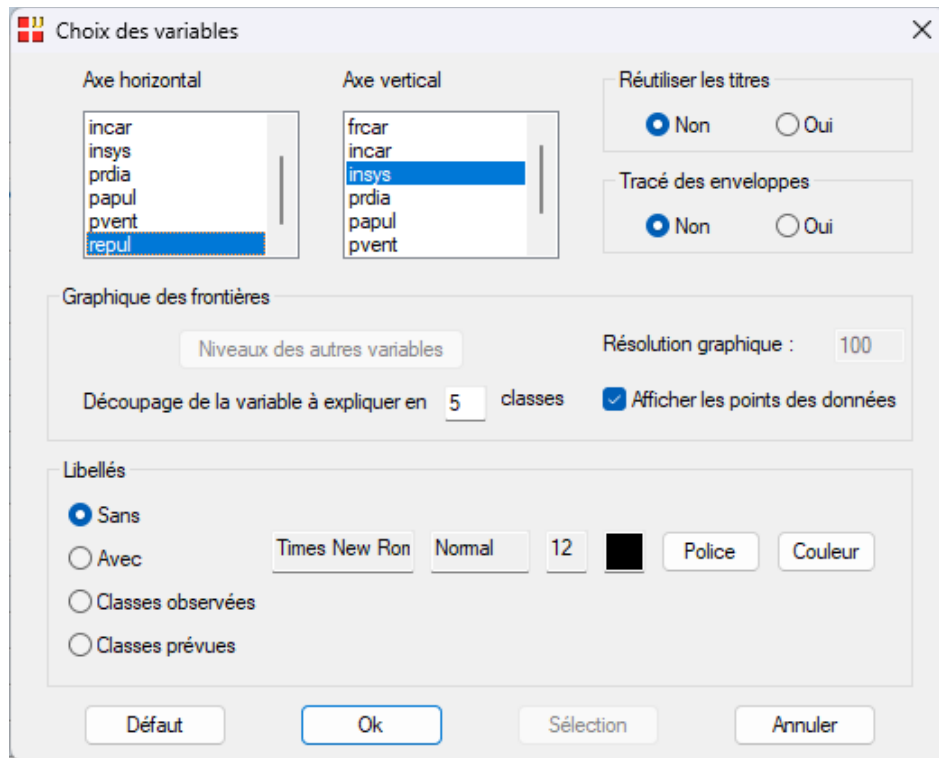
- Graphique de l'importance des variables explicatives



Les deux variables les plus importantes sont *repul* (résistance pulmonaire) et *insys* (index systolique).

- Graphique Nuage de points (apprentissage)

Cette option permet d'afficher un nuage des observations pour deux variables sélectionnées. Il est possible d'afficher les points avec ou sans libellés, avec les codes des classes observées ou avec les codes des classes prévues. Dans le graphique, la barre d'outils permet de modifier les variables affichées.



- Graphique Nuage de points (validation)

Non disponible dans cet exemple car il n'y a pas de jeu de validation.

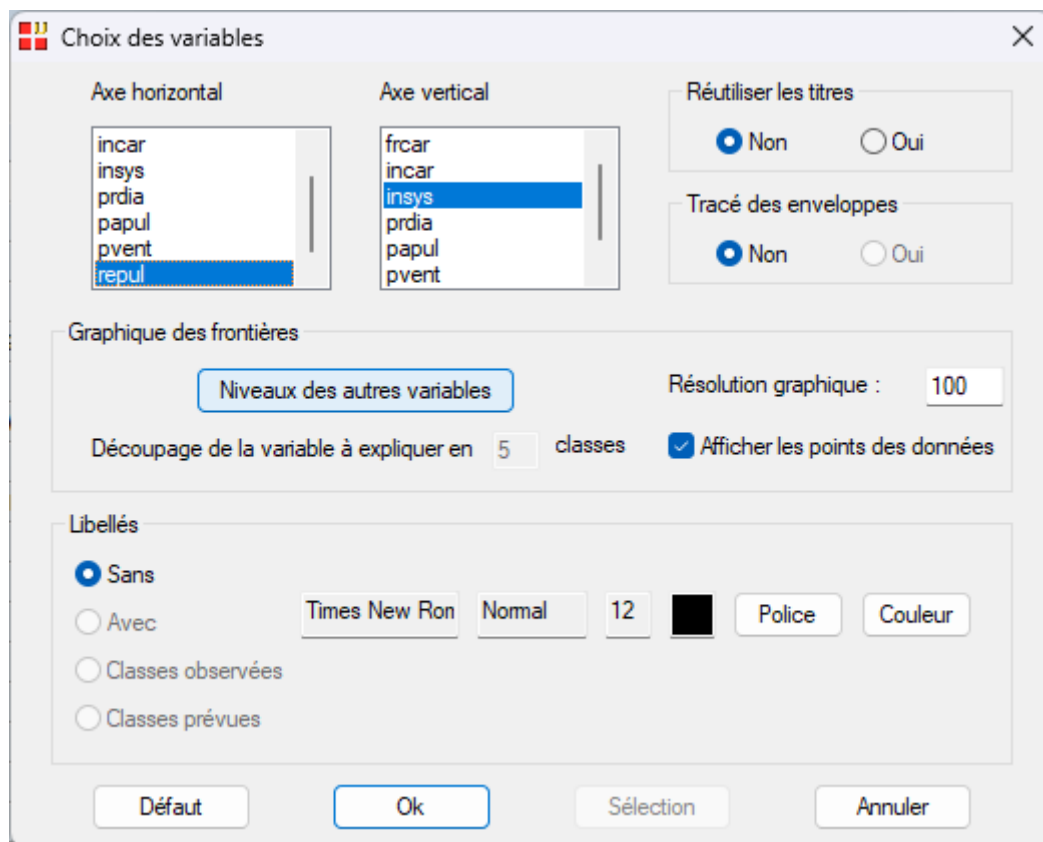
- Graphique Nuage de points (prévision)

Non disponible dans cet exemple car il n'y a pas de jeu de prévision.

- Graphique des frontières (apprentissage)

Cette option permet d'afficher un nuage des observations pour deux variables explicatives quantitatives sélectionnées et de préciser les niveaux auxquels les autres variables sont maintenues.

Par défaut les deux variables automatiquement sélectionnées sont les deux variables explicatives quantitatives les plus importantes.



Pour préciser les niveaux des autres variables, il faut cliquer sur le bouton 'Niveaux des autres variables' dans la boîte de dialogue 'Choix des variables'.

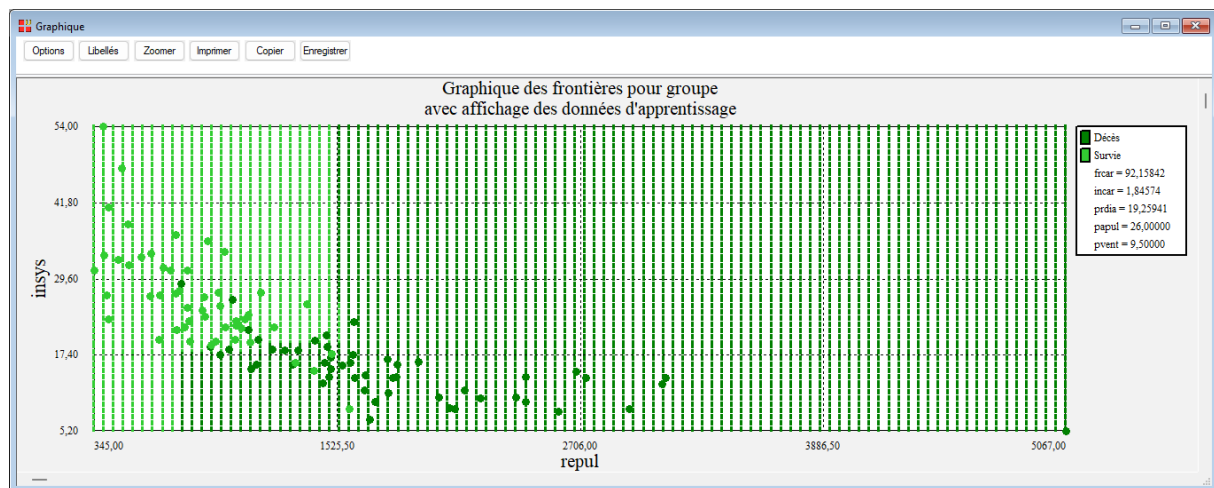
Par défaut les autres variables quantitatives sont maintenues aux valeurs moyennes des données du jeu d'apprentissage.

L'option 'Résolution graphique' permet de préciser le nombre d'évaluations de chacune des variables.

	Variable	Niveau
1	frcar	92,15842
2	incar	1,84574
3	prdia	19,25941
4	papul	26,00000
5	pvent	9,50000

Niveaux

Ok



- Graphique des frontières (validation)

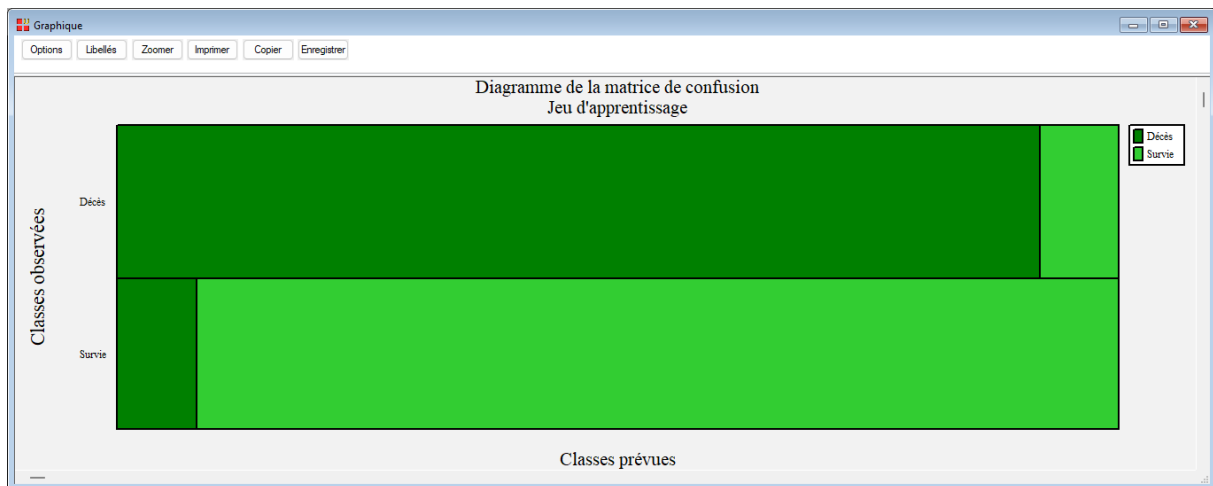
Non disponible dans cet exemple car il n'y a pas de jeu de validation.

- Diagramme de la matrice de confusion (apprentissage)

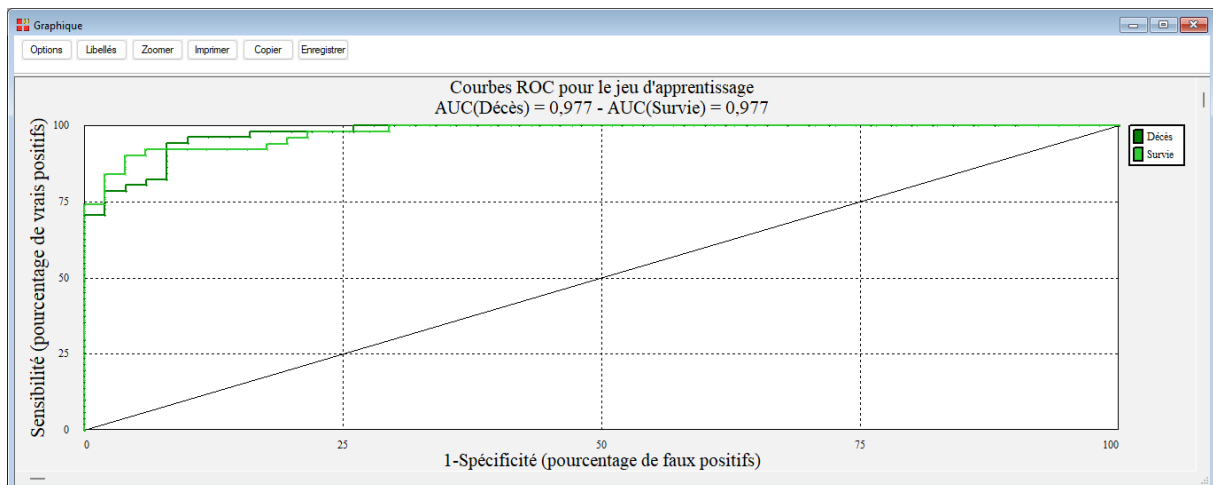
Ce diagramme permet de visualiser les résultats obtenus dans le tableau 'Matrice de confusion'.

- Diagramme de la matrice de confusion (validation)

Non disponible dans cet exemple car il n'y a pas de jeu de validation.



- Graphique de la courbe ROC (apprentissage)



L'aire sous la courbe (AUC) nous indique l'efficacité du modèle construit.

Plus la valeur de cette aire est élevée, meilleures sont les performances du modèle pour faire la distinction entre les classes *Survie* et *Décès*.

- Graphique de la courbe ROC (validation)

Non disponible dans cet exemple car il n'y a pas de jeu de validation.

Exemple 2 : Classement multi-classes - Fichier Iris3

Nous utiliserons le fichier IRIS3 pour illustrer cette procédure. Ce fichier contient pour 150 iris de trois espèces différentes les mesures des quatre caractéristiques suivantes exprimées en millimètres : longueur du sépale, largeur du sépale, longueur du pétale, largeur du pétale.

Les trois espèces sont : Setosa, Versicolor et Virginica.

Ce fichier contient 6 iris pour lesquels les classes d'appartenance sont inconnues. Ils définiront le jeu de prévision.

Sélectionnons la variable *codesp2* comme variable à expliquer, les variables *lonsepal* à *larpetal* comme variables explicatives, la variable *mesures* pour les libellés des variables explicatives et la variable *numiris* pour les libellés des observations.

Renforcement du gradient

type
lonsepal
larsepal
lonpetal
larpetal
codesp1
codesp2
numiris
mesures
nomesp

Méthode :
☒ Classement ☐ Régression

Nombre maximum d'arbres : 100
Profondeur maximale des arbres : 3
Taille minimum des noeuds terminaux : 10
Proportion des observations : 0,5
Taux d'apprentissage : 0,1
Nombre de validations croisées : 5
Racine aléatoire : 16807

Variable à expliquer :
codesp2

Variables explicatives quantitatives :
lonsepal
larsepal
lonpetal
larpetal

Variables explicatives qualitatives :

(Poids des observations :)

(Libellés des variables quantitatives :)
mesures

(Libellés modalités des variables qualitatives :)

(Libellés des observations :)
numiris

Ok Annuler Sélection Supprimer Aide

Cliquons sur le bouton *Sélection* pour définir le jeu d'apprentissage :

Liaison	Variable	Relation	Valeur ou variable
Et	larpetal	=	codesp1
Et non	larsepal	<>	codesp2
Ou	lonpetal	<	larpetal
Ou non	lonsepal	<=	larsepal
	mesures	>	lonpetal
	nomesp	>=	lonsepal
	numiris	début	mesures
	type		nomesp

☒ Trier alphabétiquement les noms des variables

Ok Annuler Ajouter Aide

Sélection

La sélection définie comporte 120 éléments.

OK

120 iris définissent le jeu d'apprentissage.

Cliquons sur le bouton OK.

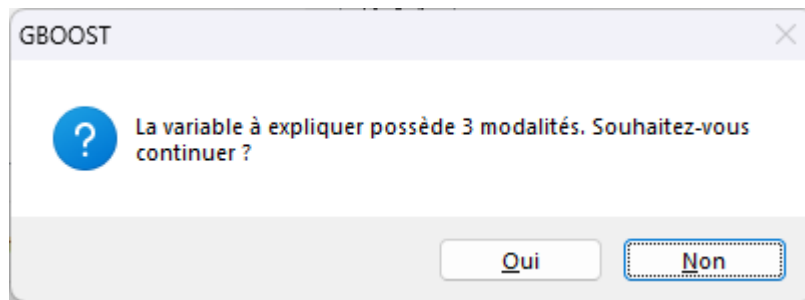
Données manquantes

Les lignes ayant des données manquantes pour la variable à expliquer définissent le jeu de prévision.

OK

Un message nous informe que les lignes ayant des données manquantes pour la variable à expliquer définissent le jeu de prévision.

Cliquons sur le bouton OK.



Confirmons que la méthode de classement utilise une variable à 3 modalités.

Après quelques instants, la fenêtre suivante s'affiche :

The screenshot shows a window titled "Rapports et Graphiques" with a sidebar on the left containing a tree view of report sections. The main area displays a table with 8 columns and 21 rows. The table contains various parameters and results of the GBOOST analysis.

	1	2	3	4	5	6	7	8
1								
2	(C) UNWIN version 10.6.0							
3								
4	DATE : 19/04/2026							
5	ORDINATEUR : LAPTOP-LEG8L077							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNÉES OUVERT(S) : IRIIS3.SGD							
8								
9	RESULTATS DU CLASSEMENT PAR RENFORCEMENT DU GRADIENT							
10								
11	Sélection :							
12	Et type = A							
13								
14	Jeu d'apprentissage : 120 observations							
15	Jeu de validation : 24 observations							
16	Jeu de prévision : 6 observations							
17								
18	Poids des observations :							
19	tous égaux à 1							
20								
21	Variable à expliquer :							

Le premier tableau, affiché ci-dessus, indique que le jeu d'apprentissage est composé de 120 observations, le jeu de validation de 24 observations et le jeu de prévision de 6 observations.

Il rappelle les modalités de la variable à expliquer, les variables explicatives utilisées et les paramètres de l'analyse.

Le deuxième tableau indique pour chaque itération (nombre d'arbres ajustés) l'évolution de la fonction de perte (log-vraisemblance) évaluée en utilisant soit les données du jeu d'apprentissage, soit par validation croisée, soit avec la proportion des données non utilisées (OOB = Out Of Bag) du jeu d'apprentissage.

Il précise également le nombre optimal d'arbres en se basant sur la validation croisée. Il est ici de 34.

Ce nombre optimal sera utilisé pour la suite des calculs.

Rapports et Graphiques

Rapport GBOOST

- Erreurs et améliorations OOB
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Observés, prévus, probabilités (valid)
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Erreurs (apprentissage, validation croisée) et améliorations OOB							
3								
4	Nombre optimal d'arbres basé sur la validation croisée : 34							
5								
6								
7	Nombre d'arbres	Apprentissage	Validation croisée	Améliorations OOB				
8	1	1,09861	1,09861	0,39056				
9	2	0,84266	0,89033	0,20425				
10	3	0,69763	0,72526	0,20494				
11	4	0,56257	0,58037	0,15956				
12	5	0,45907	0,47744	0,12394				
13	6	0,37543	0,40092	0,06075				
14	7	0,32481	0,33771	0,07878				
15	8	0,27145	0,29037	0,05730				
16	9	0,23034	0,24672	0,05373				
17	10	0,19415	0,21129	0,03909				
18	11	0,16717	0,18847	0,02970				
19	12	0,14317	0,16300	0,02595				
20	13	0,12381	0,14407	0,01359				
21	14	0,11112	0,12889	0,01829				

Rapport Explorateur /

Le troisième tableau affiche l'importance des variables explicatives dans le modèle.

Rapports et Graphiques

Rapport GBOOST

- Erreurs et améliorations OOB
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Observés, prévus, probabilités (valid)
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Importance des variables explicatives							
3								
4	(basée sur l'impact sur la performance du modèle suite à la permutation aléatoire des données du jeu d'apprentissage dans chacune des variables explicatives)							
5								
6								
7	Importance positive - la variable est utile au modèle							
8	Importance négative - la variable dégrade le modèle							
9								
10								
11		Importance						
12	lonpetal	60,23678						
13	larpetal	34,90408						
14	larsepal	3,73398						
15	lonsepal	1,12516						
16								
17								
18								
19								
20								
21								

Rapport Explorateur /

Le quatrième tableau affiche pour chaque observation du jeu d'apprentissage les valeurs observées et prévues de la variable à expliquer ainsi que les probabilités d'affectation à chacune des classes.

Rapports et Graphiques

Rapport GBOOST

- Erreurs et améliorations OOB
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Observés, prévus, probabilités (valid)
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Jeu d'apprentissage - classes observées, classes prévues et probabilités des classes							
3								
4	Classes :							
5								
6	1 = Setosa							
7	2 = Versicolor							
8	3 = Virginica							
9								
10								
11	Observation	Classe observée	Classe prévue	Setosa	Versicolor	Virginica		
12	1	1	1	0,99895	0,00061	0,00044		
13	2	1	1	0,99867	0,00051	0,00083		
14	4	1	1	0,99893	0,00047	0,00060		
15	5	1	1	0,99904	0,00052	0,00044		
16	6	1	1	0,99872	0,00088	0,00039		
17	7	1	1	0,99907	0,00057	0,00037		
18	8	1	1	0,99897	0,00072	0,00031		
19	9	1	1	0,99828	0,00089	0,00082		
20	11	1	1	0,99895	0,00062	0,00044		
21	12	1	1	0,99902	0,00057	0,00041		

Rapport Explorateur /

Le tableau suivant affiche la matrice de confusion pour le jeu d'apprentissage.

Rapports et Graphiques

Rapport GBOOST

- Erreurs et améliorations OOB
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)**
- Observés, prévus, probabilités (valid)
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Matrice de confusion pour le jeu d'apprentissage							
3								
4	Basée sur le tableau des prévisions							
5								
6	En lignes, les classes observées							
7	En colonnes, les classes prévues							
8								
9	Pourcentage de mal classés : 0,00 %							
10	Pourcentage de bien classés (exactitude) : 100,00 %							
11								
12	Précision = $VP / (VP + FP)$							
13	Rappel = $VP / (VP + FN)$							
14	Score F1 = $2 \times (\text{Précision} \times \text{Rappel}) / (\text{Précision} + \text{Rappel})$							
15								
16								
17	Observé \ Prévu	Taille	Setosa	Versicolor	Virginica	Précision	Rappel	Score F
18	Setosa	43	43	0	0	1	1	
19	Versicolor	38	0	38	0	1	1	
20	Virginica	39	0	0	39	1	1	
21								

Rapport Explorateur /

Les deux tableaux suivants affichent ces mêmes informations pour le jeu de validation.

Rapports et Graphiques

Rapport GBOOST

- Erreurs et améliorations OOB
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Observés, prévus, probabilités (valid)**
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Jeu de validation - classes observées, classes prévues et probabilités des classes							
3								
4	Classes :							
5								
6	1 = Setosa							
7	2 = Versicolor							
8	3 = Virginica							
9								
10								
11	Observation	Classe observée	Classe prévue	Setosa	Versicolor	Virginica		
12	10	1	1	0,99693	0,0047	0,00060		
13	25	1	1	0,99678	0,00082	0,00041		
14	41	1	1	0,99908	0,00052	0,00039		
15	48	1	1	0,99903	0,00057	0,00041		
16	49	1	1	0,99895	0,00062	0,00044		
17	54	2	2	0,00098	0,99667	0,00235		
18	55	2	2	0,00145	0,99262	0,00593		
19	57	2	2	0,00245	0,98678	0,01079		
20	68	2	2	0,00099	0,99632	0,00269		
21	78 (*)	2	3	0,00087	0,00353	0,99561		

Rapport Explorateur /

Rapports et Graphiques

Rapport GBOOST

- Erreurs et améliorations OOB
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Matrice de confusion (valid)**
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Matrice de confusion pour le jeu de validation							
3								
4	Basée sur le tableau des prévisions							
5								
6	En lignes, les classes observées							
7	En colonnes, les classes prévues							
8								
9	Pourcentage de mal classés : 4,167 %							
10	Pourcentage de bien classés (exactitude) : 95,833 %							
11								
12	Précision = $VP / (VP + FP)$							
13	Rappel = $VP / (VP + FN)$							
14	Score F1 = $2 \times (\text{Précision} \times \text{Rappel}) / (\text{Précision} + \text{Rappel})$							
15								
16								
17	Observé \ Prévu	Taille	Setosa	Versicolor	Virginica	Précision	Rappel	Score F
18	Setosa	5	5	0	0	1,0	1,0	1,0000
19	Versicolor	10	0	9	1	1,0	0,9	0,9473
20	Virginica	9	0	0	9	0,9	1,0	0,9473
21								

Rapport Explorateur /

Le dernier tableau affiche les résultats pour le jeu de prévision.

Rapports et Graphiques

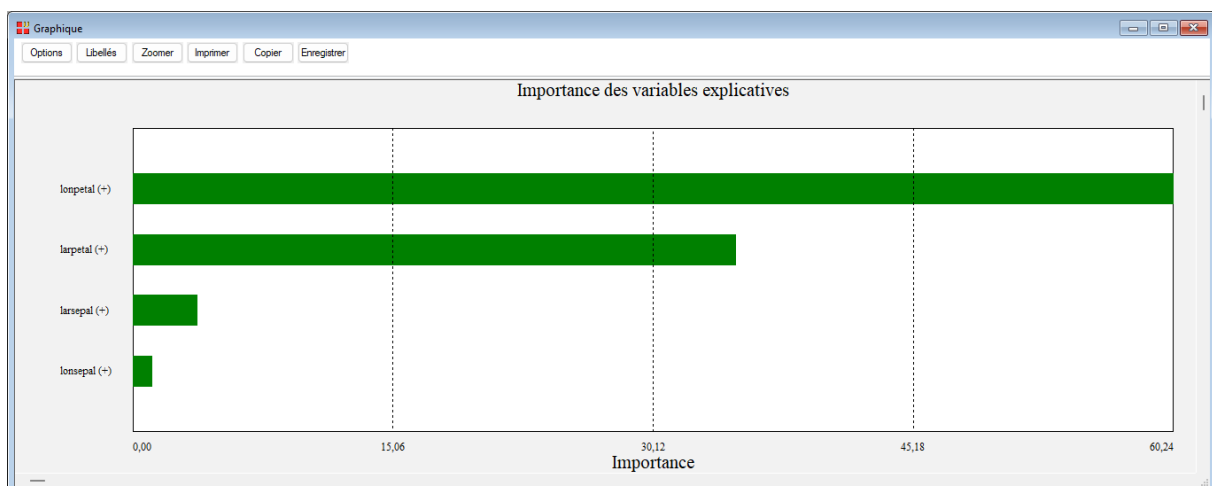
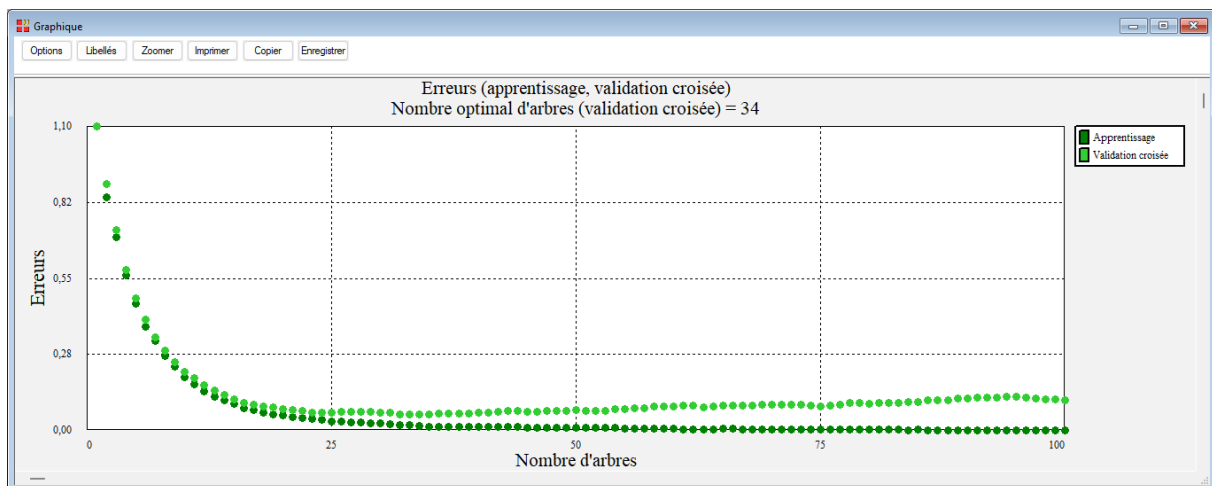
Rapport GBOOST

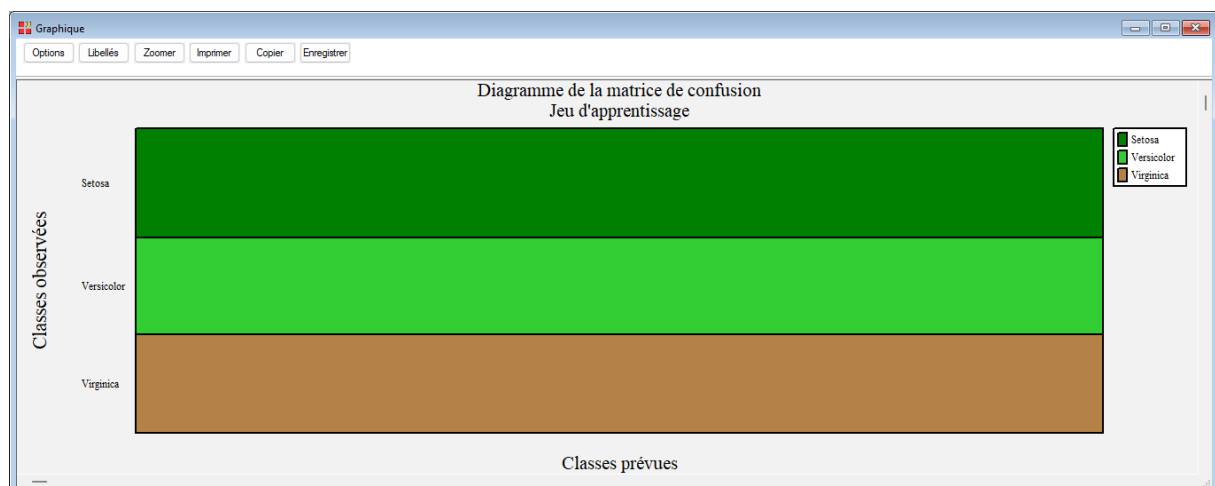
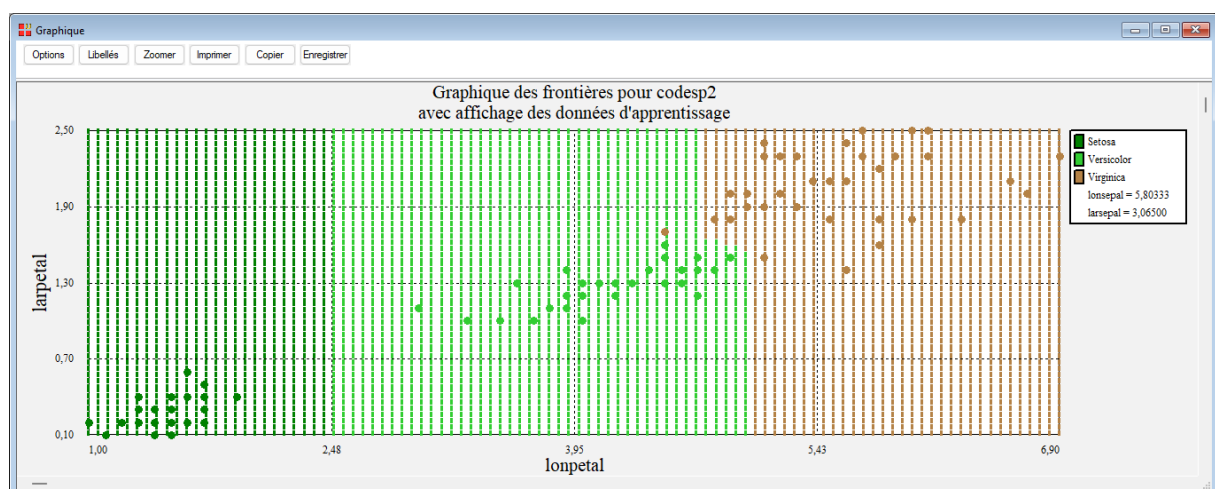
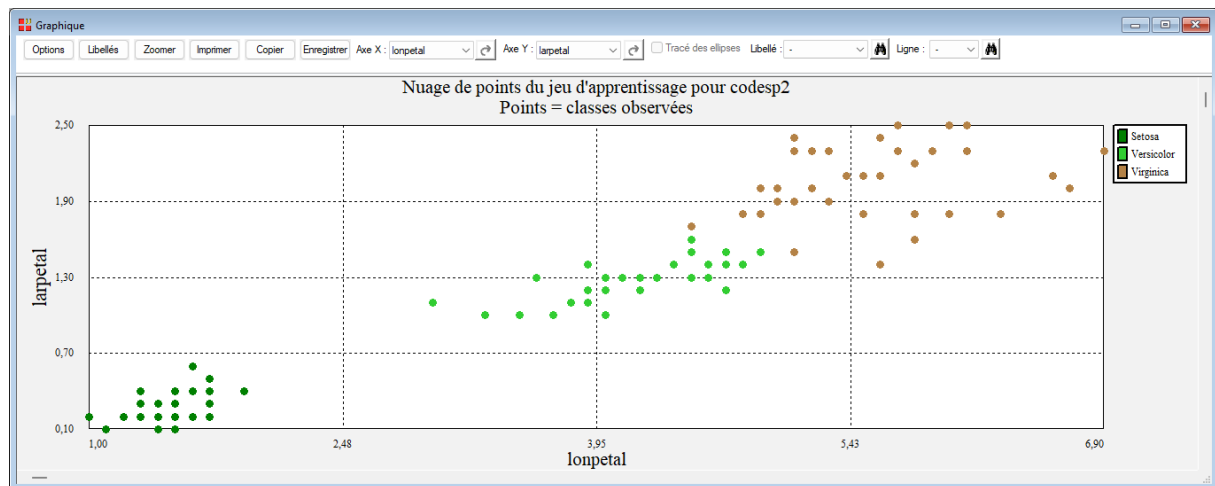
- Erreurs et améliorations OOB
- Importance des variables (app)
- Observés, prévus, probabilités (app)
- Matrice de confusion (app)
- Observés, prévus, probabilités (valid)
- Matrice de confusion (valid)
- Prévus, probabilités (prévision)

	1	2	3	4	5	6	7	8
1								
2	Jeu de prévision - classes prévues et probabilités des classes							
3								
4	Classes :							
5								
6	1 = Setosa							
7	2 = Versicolor							
8	3 = Virginica							
9								
10								
11	Observation	Classe prévue	Setosa	Versicolor	Virginica			
12	3	1	0,99903	0,00057	0,00041			
13	36	1	0,99897	0,00072	0,00031			
14	62	2	0,00106	0,99606	0,00287			
15	84	3	0,00081	0,00245	0,99674			
16	104	3	0,00034	0,00049	0,99917			
17	125	3	0,00145	0,00521	0,99334			
18								
19								
20								
21								

Rapport Explorateur

Voici quelques graphiques issus de cette analyse.





Exemple 3 : Régression - Fichier Boston

Ce fichier contient des informations collectées par le « U.S Census Service » concernant la valeur des habitations dans l'agglomération de Boston (source : <http://lib.stat.cmu.edu/datasets/boston>).

Les caractéristiques contenues dans le fichier sont les suivantes :

crim	taux de criminalité par habitant par ville
zn	proportion de terrains résidentiels zonés pour des lots de plus de 25 000 pieds carrés.
indus	proportion d'acres non commerciales par ville
chas	indicatrice de proximité à la Charles River (1 = proche, 0 = sinon)
nox	concentration d'oxydes nitriques (partie pour 10 millions)
rm	nombre moyen de pièces par habitation
age	proportion des habitations construites avant 1940
dis	distance pondérée à cinq lieux d'emplois de Boston
rad	indice d'accessibilité aux autoroutes
tax	taux d'imposition foncière par \$ 10.000
ptratio	ratio élèves / enseignants par ville
black	$1000(Bk-0,63)^2$ où Bk = proportion de noirs par ville
lstat	% statut inférieur de la population
medv	valeur médiane de l'habitation (en milliers de \$)
type	indique si l'observation est utilisée pour l'apprentissage (A) ou la validation (V)

Le but est de prévoir la variable *medv* en utilisant ces données.

Cliquons sur l'icône GBOOST dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-après.

La variable à expliquer est *medv* et les variables explicatives sont *crim*, *zn*, *indus*, *chas*, *nox*, *rm*, *age*, *dis*, *rad*, *tax*, *ptratio*, *black*, *lstat*.

Cliquons sur le bouton 'Sélection' pour sélectionner les données du jeu d'apprentissage (type=A).

391 observations sont sélectionnées, 105 observations définissent le jeu de validation et 10 le jeu de prévision.

Cliquons sur Ok.

Renforcement du gradient

obs
crim
zn
indus
chas
nox
rm
age
dis
rad
tax
ptratio
black
lstat
medv
type

Méthode :
☐ Classement
☒ Régression

Nombre maximum d'arbres : 100

Profondeur maximale des arbres : 3

Taille minimum des noeuds terminaux : 10

Proportion des observations : 0,5

Taux d'apprentissage : 0,1

Nombre de validations croisées : 5

Racine aléatoire : 16807

Variable à expliquer : medv

Variables explicatives quantitatives :
crim
zn
indus
chas
nox
rm
age
dis

Variables explicatives qualitatives :

(Poids des observations :)

(Libellés des variables quantitatives :)

(Libellés modalités des variables qualitatives :)

(Libellés des observations :)
obs

Ok Annuler Sélection Supprimer Aide

Définition de la sélection

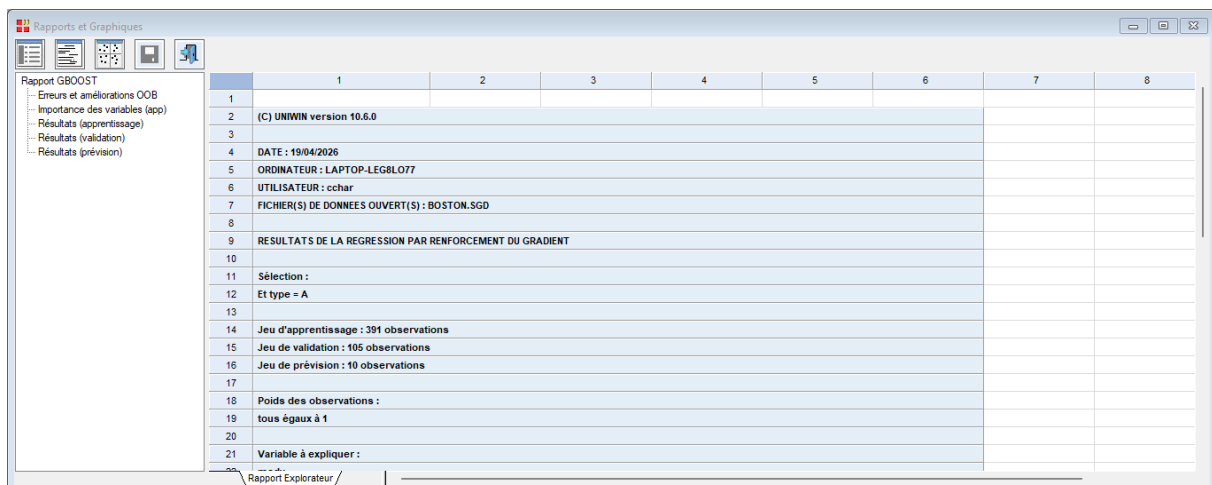
Et type = A

Liaison	Variable	Relation	Valeur ou variable
Et	nox	=	age
Et non	obs	<>	black
Ou	ptratio	<	chas
Ou non	rad	<=	crim
	rm	>	dis
	tax	>=	indus
	type	début	lstat
	zn		medv

☒ Trier alphabétiquement les noms des variables

Ok Annuler Ajouter Aide

Après quelques instants, le tableau suivant s'affiche :

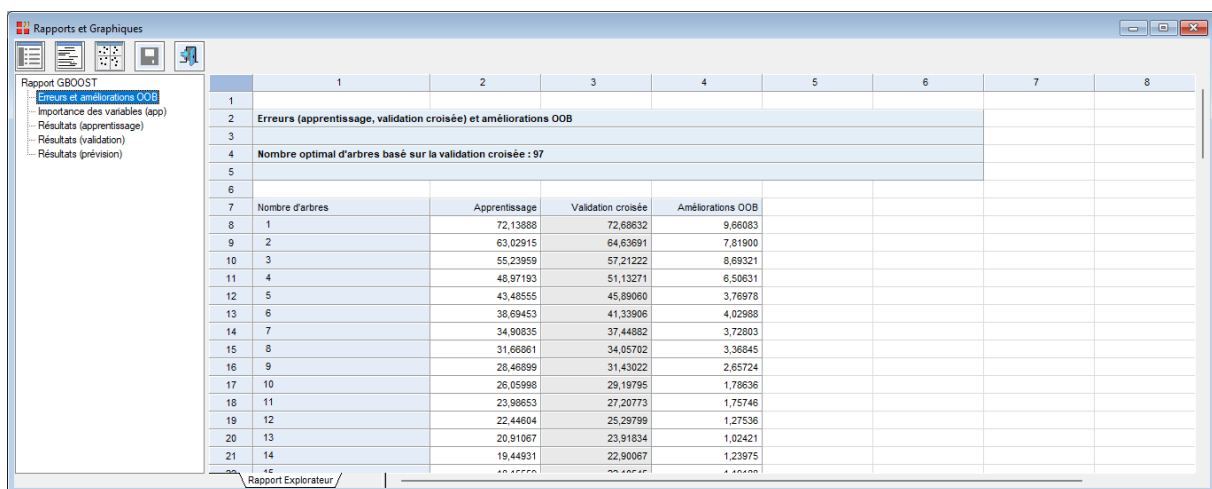


	1	2	3	4	5	6	7	8
1								
2	(C) UNIWIN version 10.6.0							
3								
4	DATE : 19/04/2026							
5	ORDINATEUR : LAPTOP-LEG8LO77							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNEES OUVERT(S) : BOSTON.SGD							
8								
9	RESULTATS DE LA REGRESSION PAR RENFORCEMENT DU GRADIENT							
10								
11	Sélection :							
12	Et type = A							
13								
14	Jeu d'apprentissage : 391 observations							
15	Jeu de validation : 105 observations							
16	Jeu de prévision : 10 observations							
17								
18	Poids des observations :							
19	tous égaux à 1							
20								
21	Variable à expliquer :							

Le deuxième tableau indique pour chaque itération (nombre d'arbres ajustés) l'évolution de la fonction de perte (log-vraisemblance) évaluée en utilisant soit les données du jeu d'apprentissage, soit par validation croisée, soit avec la proportion des données non utilisées (OOB = Out Of Bag) du jeu d'apprentissage.

Il précise également le nombre optimal d'arbres en se basant sur la validation croisée. Il est ici de 97.

Ce nombre optimal sera utilisé pour la suite des calculs.



	1	2	3	4	5	6	7	8
1								
2	Erreurs (apprentissage, validation croisée) et améliorations OOB							
3								
4	Nombre optimal d'arbres basé sur la validation croisée : 97							
5								
6								
7	Nombre d'arbres	Apprentissage	Validation croisée	Améliorations OOB				
8	1	72,13888	72,68632	9,66083				
9	2	63,02915	64,63691	7,81900				
10	3	55,23959	57,21222	6,69321				
11	4	48,97193	51,13271	6,50631				
12	5	43,48555	45,89060	3,76978				
13	6	38,69453	41,33906	4,02988				
14	7	34,90835	37,44882	3,72803				
15	8	31,66861	34,05702	3,36645				
16	9	28,46889	31,43022	2,65724				
17	10	26,05998	29,19795	1,78636				
18	11	23,98653	27,20773	1,75746				
19	12	22,44604	25,29799	1,27536				
20	13	20,91067	23,91834	1,02421				
21	14	19,44931	22,90067	1,23975				

Le troisième tableau affiche l'importance des variables explicatives dans le modèle.

Les deux tableaux suivants affichent les résultats (valeurs observées, prévues et résidus) pour les jeux d'apprentissage, de validation.

Enfin, le dernier tableau affiche les résultats (valeurs prévues) pour le jeu de prévision.

Rapports et Graphiques								
Rapport GBOOST								
Erreurs et améliorations OOB								
Importance des variables (app)								
Résultats (apprentissage)								
Résultats (validation)								
Résultats (prévision)								
1	1	2	3	4	5	6	7	8
2	Importance des variables explicatives							
3								
4	(basée sur l'impact sur la performance du modèle suite à la permutation aléatoire							
5	des données du jeu d'apprentissage dans chacune des variables explicatives)							
6								
7	Importance positive = la variable est utile au modèle							
8	Importance négative = la variable dégrade le modèle							
9								
10								
11								
12	lstat	Importance						
13	rm	46,23115						
14	dis	30,85517						
15	nox	5,61155						
16	crim	4,88379						
17	ptratio	3,88893						
18	black	3,08055						
19	tax	1,56392						
20	age	1,38268						
21	rad	1,21800						
22	rad	0,45742						

Rapports et Graphiques								
Rapport GBOOST								
Erreurs et améliorations OOB								
Importance des variables (app)								
Résultats (apprentissage)								
Résultats (validation)								
Résultats (prévision)								
1	1	2	3	4	5	6	7	8
2	Résultats pour le jeu d'apprentissage							
3								
4	R-carré = 92,815 %							
5								
6								
7	Observation	Observé	Prévu	Résidu				
8	o1	24.0	26,70585	-2,70585				
9	o2	21.6	22,34871	-0,74871				
10	o3	34.7	35,76338	-1,06338				
11	o4	33.4	37,02333	-3,62333				
12	o5	36.2	33,28940	2,91060				
13	o6	28.7	26,60418	2,09582				
14	o7	22.9	22,01723	0,88277				
15	o8	27.1	20,55194	6,54806				
16	o11	15.0	17,08284	-2,08284				
17	o12	18.9	20,18705	-1,28705				
18	o13	21.7	19,74113	1,95887				
19	o14	20.4	19,66259	0,73741				
20	o15	18.2	18,77813	-0,57813				
21	o16	19.9	19,67382	0,22618				

Rapports et Graphiques								
Rapport GBOOST								
Erreurs et améliorations OOB								
Importance des variables (app)								
Résultats (apprentissage)								
Résultats (validation)								
Résultats (prévision)								
1	1	2	3	4	5	6	7	8
2	Résultats pour le jeu de validation							
3								
4	R-carré = 86,008 %							
5								
6								
7	Observation	Observé	Prévu	Résidu				
8	o10	18.9	17,93063	0,96937				
9	o18	17.5	17,33670	0,16330				
10	o30	21.0	22,21365	-1,21365				
11	o31	12.7	13,52087	-0,82087				
12	o33	13.2	13,08055	0,11945				
13	o43	25.3	25,43147	-0,13147				
14	o48	16.6	18,34911	-1,74911				
15	o49	14.4	16,48209	-2,08209				
16	o57	24.7	22,22478	2,47522				
17	o66	23.5	26,16330	-2,66330				
18	o68	22.0	20,37712	1,62288				
19	o73	22.8	22,79230	0,00770				
20	o75	24.1	24,22698	-0,12698				
21	o77	20.0	21,72298	-1,72298				

Rapports et Graphiques								
Rapport GBOOST								
Erreurs et améliorations OOB								
Importance des variables (app)								
Résultats (apprentissage)								
Résultats (validation)								
Résultats (prévision)								
Résultats pour le jeu de prévision								
1								
2								
3								
4								
5	Observation	Prévu						
6	o9	16,91781						
7	o25	15,45287						
8	o60	20,61045						
9	o97	23,01386						
10	o141	14,97834						
11	o174	25,23837						
12	o202	24,08878						
13	o238	39,58829						
14	o351	20,98747						
15	o495	20,65872						
16								
17								
18								
19								
20								
21								

Les graphiques disponibles pour un modèle de régression sont :

Graphiques

☒ Graphique des erreurs (apprentissage, validation croisée)

☐ Graphique de l'importance des variables explicatives

☐ Nuage de points (apprentissage)

☐ Nuage de points (validation)

☐ Nuage de points (prévision)

☐ Graphique des frontières (apprentissage)

☐ Graphique des frontières (validation)

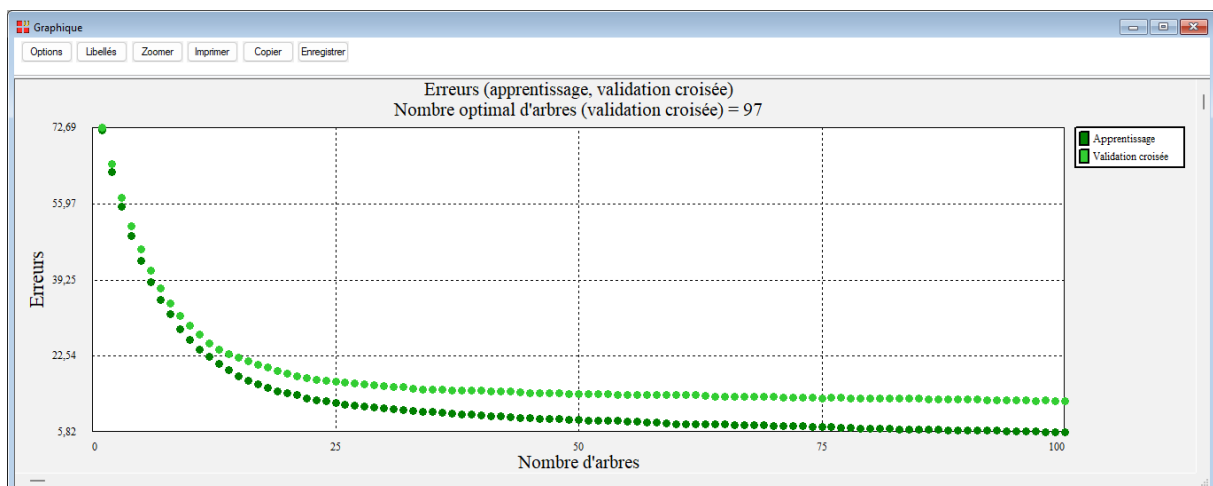
☐ Prévus versus observés (apprentissage)

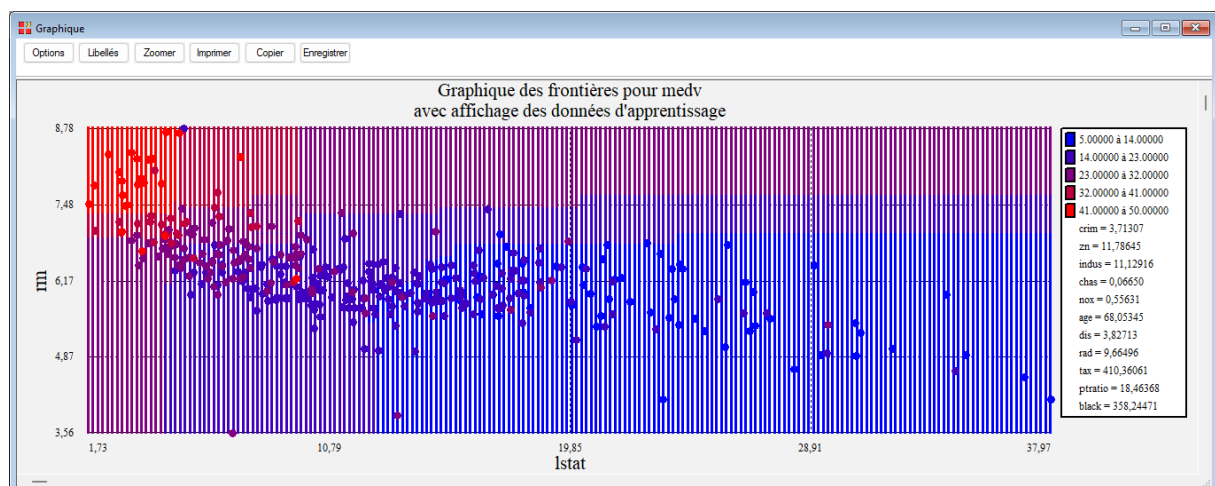
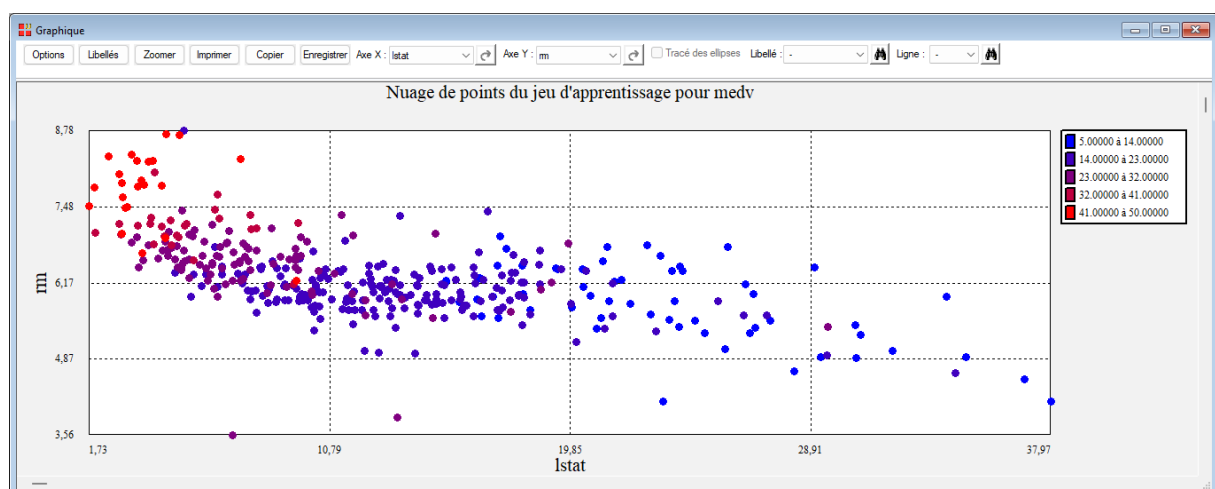
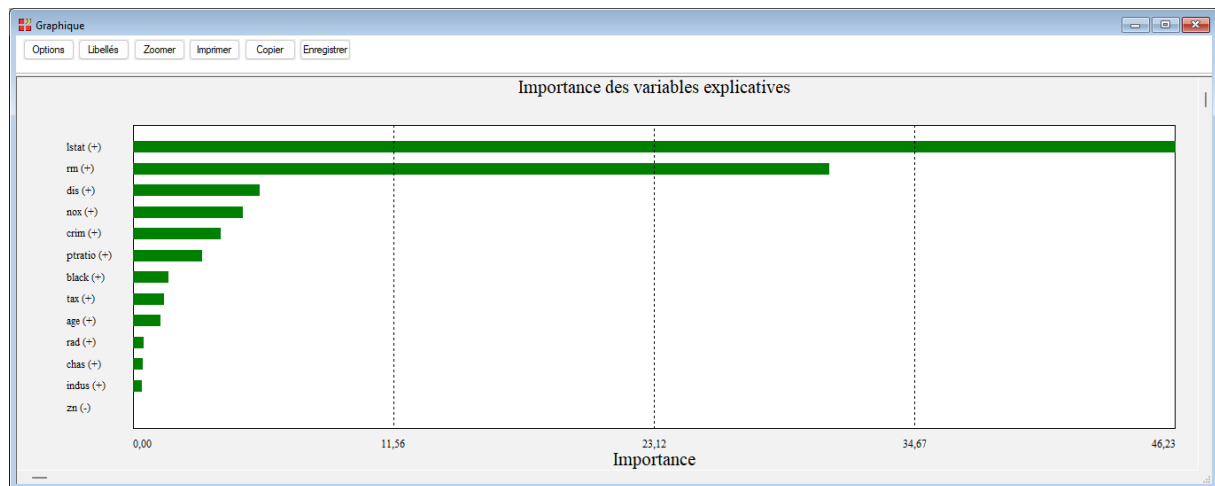
☐ Prévus versus observés (validation)

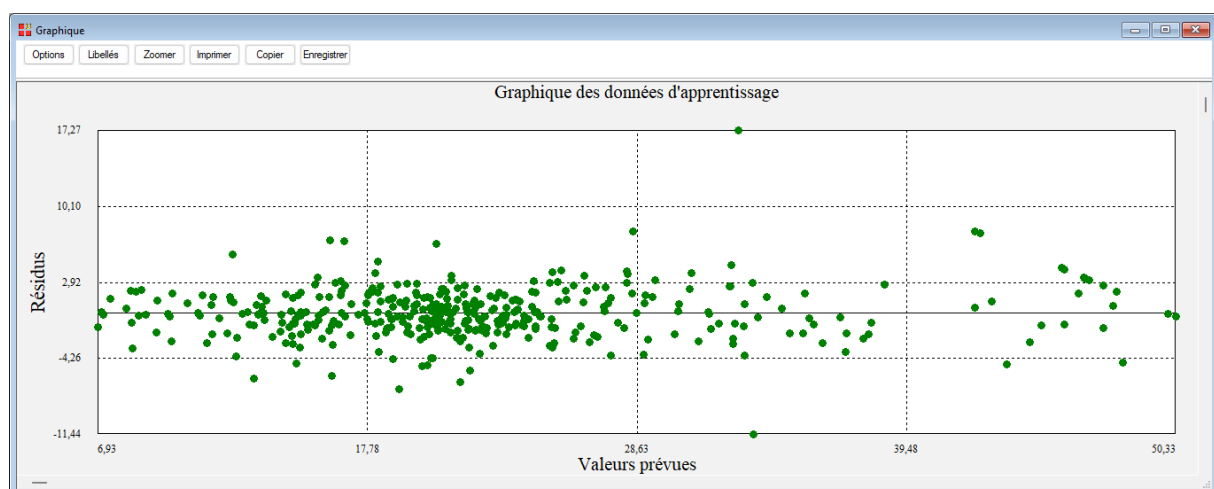
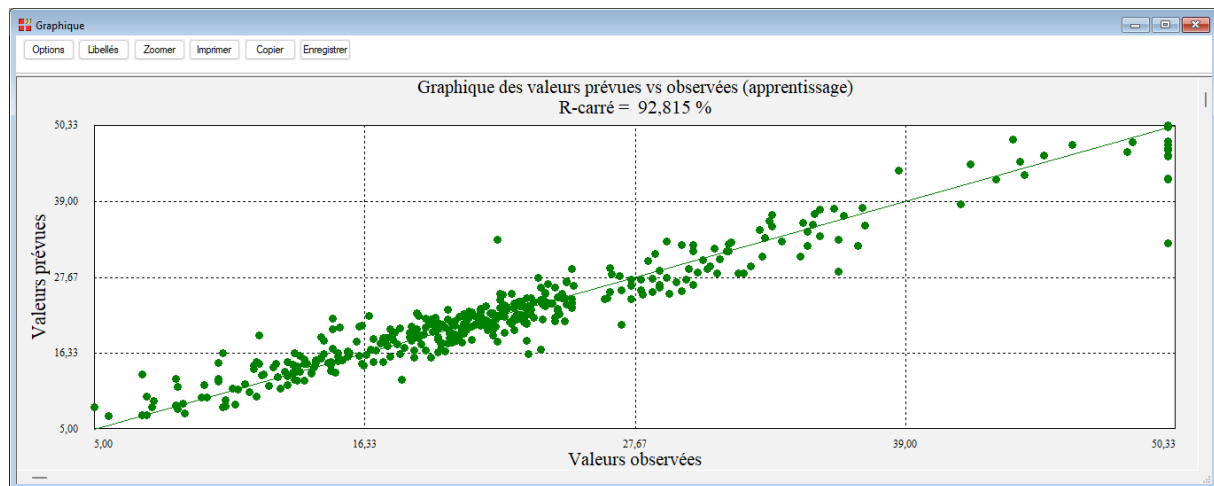
☐ Résidus versus prévus (apprentissage)

☐ Résidus versus prévus (validation)

Ok Annuler







Enregistrement des résultats

Voici la liste des variables créées par la procédure.

Variable	Contenu
erreurapp	Erreurs d'apprentissage
erreurcv	Erreurs de validation croisée
amelooob	Améliorations OOB
varexpli	Variables explicatives
impvarexpli	Importance des variables explicatives
libobsapp	Libellés des observations (apprentissage)
probapp	Probabilités (classement, apprentissage)
obsapp	Valeurs observées (apprentissage)
vpapp	Valeurs prévues (apprentissage)
redisapp	Résidus (régression, apprentissage)
libobsvalid	Libellés des observations (validation)
probvalid	Probabilités (classement, validation)
obsbvalid	Valeurs observées (validation)

vpvalid	Valeurs prévues (validation)
residvalid	Résidus (régression, validation)
libobsprev	Libellés des observations (prévision)
probprev	Probabilités (prévision)
vpprev	Valeurs prévues (prévision)
seuilA	Seuils (apprentissage, classement binaire)
specificiteA	Spécificités (apprentissage, classement binaire)
sensibiliteA	Sensibilités (apprentissage, classement binaire)
aireA	Aires sous les courbes (apprentissage, classement binaire)
seuilV	Seuils (validation, classement binaire)
specificiteV	Spécificités (validation, classement binaire)
sensibiliteV	Sensibilités (validation, classement binaire)
aireV	Aires sous les courbes (validation, classement binaire)

Références

Package R 'gbm' (2026)

<https://cran.r-project.org/web/packages/gbm/gbm.pdf>

Friedman, J.H. (2001). *Greedy Function Approximation : A Gradient Boosting Machine*. Annals of Statistics, 29(5), 1189–1232.

<https://projecteuclid.org/euclid.aos/1013203451>

Generalized Boosted Models : A guide to the gbm package Greg Ridgeway, 2026

<file:///C:/Users/cchar/AppData/Local/R/win-library/4.5/gbm/doc/gbm.pdf>

Exemple 'Boston' : <http://lib.stat.cmu.edu/datasets/boston>