

UNIWIN VERSION 10.0.0

ANALYSE APRIORI

Révision : 30/06/2024

| | |
|--|----|
| Définition..... | 1 |
| Entrée des données | 2 |
| Données manquantes | 3 |
| Définitions..... | 3 |
| Exemple 1 : Fichier Crédit3 | 4 |
| L'option Rapports | 8 |
| L'option Graphiques | 11 |
| Exemple 2 : Fichier Income | 14 |
| Exemple 3 : Fichier Ventas2..... | 17 |
| Les variables créées par la procédure..... | 20 |
| Références | 20 |

Définition

L'algorithme APRIORI est un algorithme d'exploration de données conçu en 1994 par Rakesh Agrawal et Ramakrishnan Srikant dans le domaine de l'apprentissage des règles d'association.

Il sert à reconnaître des propriétés qui reviennent fréquemment dans un ensemble de données et à détecter les relations ou les associations entre des modalités spécifiques de variables catégorielles dans des fichiers de données volumineux.

Cette technique exploratoire puissante possède toute une gamme d'applications dans de nombreux domaines : analyse des tickets de caisse, campagne marketing, analyse de log Web, détection de fraudeurs, analyse de défaillances de process.

Cette procédure est basée sur le package R 'arules'.

Entrée des données

Cliquons sur l'icône APRIORI dans le ruban Décrire. La boîte de dialogue montrée ci-dessous s'affiche :

The dialog box 'Analyse APRIORI' contains the following elements:

- Colonnes des données :** A list box for selecting data columns, accompanied by a right-pointing arrow button.
- (Libellés des colonnes des données :)** A text input field for column labels, accompanied by a right-pointing arrow button.
- (Colonne cible :)** A text input field for the target column, accompanied by a right-pointing arrow button.
- Paramètres des règles**
 - Support minimal :
 - Confiance minimale :
 - Cardinal maximal :
 - Lift minimal :
 - Trier les listes par :
- Buttons:** Ok, Annuler, Sélection, Supprimer, Aide.

Cette boîte de dialogue permet de définir les colonnes des données, les libellés des colonnes des données, la colonne cible d'intérêt ainsi que les paramètres pour la recherche des règles (support relatif minimal, confiance relative minimale, cardinal maximal) et le lift minimal pour la sélection des règles construites. Le critère de tri des listes de règles peut également être précisé.

Toutes les données doivent être alphanumériques.

Les libellés des colonnes des données et la colonne cible sont optionnelles.

Données manquantes

Dans cette procédure les données manquantes sont autorisées et sont considérées comme des modalités des colonnes des données.

Définitions

Item : un produit, une réponse dans un questionnaire, le statut d'un élément de pilotage d'un process, ...

Motif : un ensemble d'items

Cardinal : nombre d'items dans un motif

Sur-motif : le motif B est un sur-motif du motif A si $\text{cardinal}(A) < \text{cardinal}(B)$ avec $A \subset B$ et donc $\text{support}(B) \leq \text{support}(A)$.

Motif fréquent : motif dont le support est supérieur ou égal au support minimal précisé.

Motif fréquent fermé : motif fréquent dont aucun de ses sur-motifs n'a un support identique.

Motif fréquent maximal : motif fréquent dont aucun de ses super-motifs n'est fréquent.

Les motifs fréquents fermés et maximaux permettent de réduire le nombre de règles redondantes.

Antécédent (A) : partie gauche d'une règle 'Si A alors C '

Conséquent (C) : partie droite d'une règle 'Si A alors C '

Support $A \Rightarrow C$: indicateur de fiabilité de la règle = probabilité conjointe (AC)

Confiance $A \Rightarrow C$: indicateur de précision de la règle = probabilité conditionnelle (C/A)

Une bonne règle doit avoir un support élevé et une confiance élevée.

Lift $A \Rightarrow C$: indicateur de pertinence de la règle, surcroît d'occurrence du conséquent quand l'antécédent est présent

$$\text{Lift} = \text{probabilité conditionnelle } (C/A) / \text{probabilité } (C)$$

Un lift supérieur à 1 indique qu'il existe un lien positif entre les items et que la règle est significative.

Exemple 1 : Fichier Crédit3

Pour illustrer ce premier exemple, nous utiliserons le fichier Crédit3, sous-ensemble du fichier Crédit présenté dans la documentation concernant le scoring.

Le fichier complet d'origine est disponible sur le site « UCI Machine Learning Repository ». Il a été conçu par le Professeur Dr. Hans Hofmann de l'Institut für Statistik und Okonometrie, Universität Hamburg

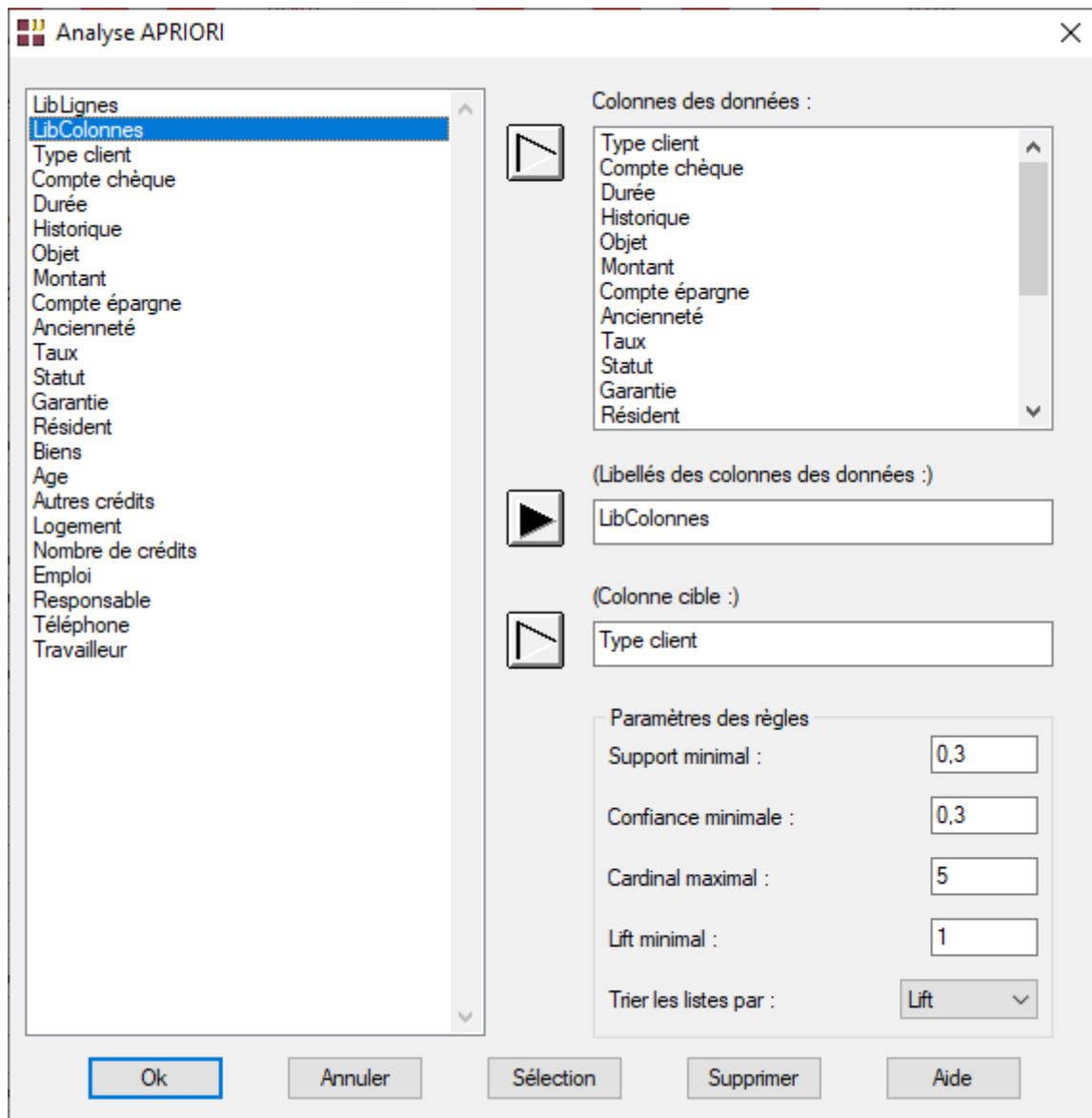
Le fichier Crédit3 contient un ensemble de 21 informations concernant 100 demandeurs d'un crédit.

Voici la liste de ces 21 informations :

| <i>Variable</i> | <i>Descriptif de la variable qualitative et de ses modalités</i> |
|-----------------|---|
| Type client | Type de client « Bon client », « Mauvais client » |
| Compte chèque | Montant sur le compte chèque en DM « 0 DM », « 1 à 200 DM », « Plus de 200 DM » « Pas de compte chèque » |
| Durée | Durée du crédit en mois « 1 mois », « 2 mois », « 3 mois », « 4 mois » |
| Historique | Historique des crédits « Pas de crédits », « Crédits terminés payés » « Crédits en cours », « Crédits payés en retard » « Compte critique » |
| Objet | Objet du crédit « Véhicule neuf », « Véhicule d'occasion » « Biens d'équipement », « Radio ou télévision » « Appareils ménagers », « Réparations » « Education », « Travail », « Autres » |
| Montant | Montant du crédit (variable continue discrétisée par quartiles) « 0 à 1366 DM », « 1367 à 2320 DM » « 2321 à 3973 DM », « 3974 DM et plus » |
| Compte épargne | Montant sur le compte d'épargne en DM « Moins de 100 DM », « 101 à 500 DM », « 501 à 1000 DM », « Plus de 1000 DM » « Sans épargne » |
| Ancienneté | Ancienneté dans l'emploi « Sans emploi », « Moins de 1 an », « 1 à 3 ans » 4 = « 4 à 6 ans », « Plus de 6 ans » |

| | |
|-------------------|---|
| Taux | Taux du crédit « 1% », « 2% », « 3% », « 4% » |
| Statut | Statut du demandeur (sexe et civilité) « H divorcé », « F divorcée ou mariée » « H célibataire », « H marié » |
| Garantie | Garantie, autres débiteurs « Aucune », « Codemandeur » « Caution » |
| Résident | Nombre d'années de résidence « 1 an », « 2 ans », « 3 ans », « 4 ans » |
| Biens | Biens possédés « Immobilier », « Epargne, Assurance » « Véhicule », « Sans biens » |
| Age | Age du demandeur ((variable continue discrétisée par quartiles) « Moins de 27 ans », « 27 à 32 ans » « 33 à 41 ans », « Plus de 41 ans » |
| Autres crédits | Autres crédits en cours « Bancaires », 2 = « Magasins » « Aucun » |
| Logement | Logement « Locataire », « Propriétaire » « Logement gratuit » |
| Nombre de crédits | Nombre de crédits en cours « 1 crédit », « 2 crédits », « 3 crédits » « 4 crédits » |
| Emploi | Type d'emploi « Sans », « Non qualifié » « Employé », « Direction » |
| Responsable | Nombre de personnes responsables du crédit « Un », « Deux » |
| Téléphone | Téléphone « Sans », « Avec » |
| Travailleur | Type de travailleur « Etranger », « National » |

Renseignons la boîte de dialogue comme montré ci-dessous.

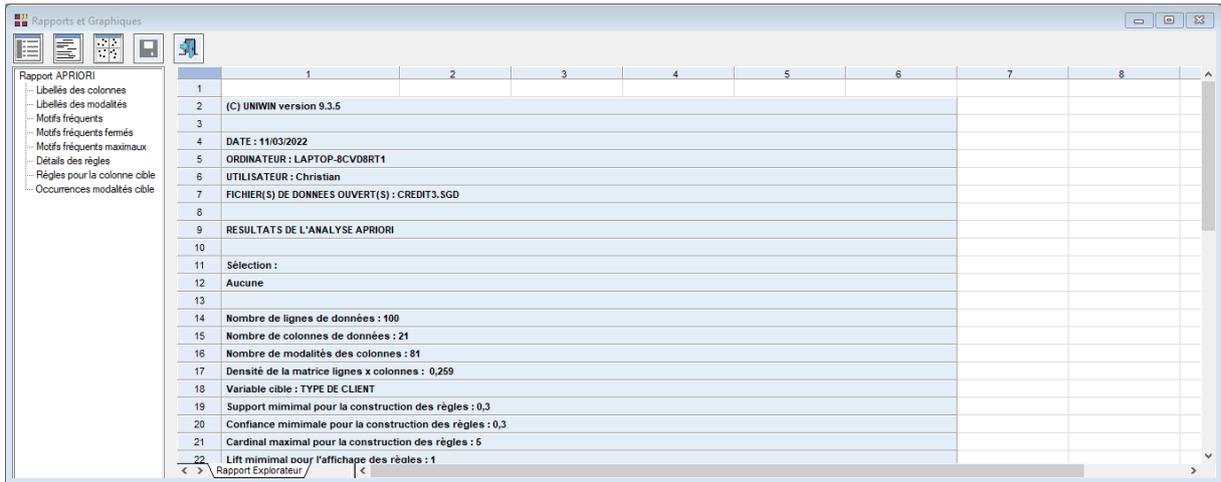


Sélectionnons les variables 'Type de client' à 'Travailleur' comme colonnes des données, la variable LibColonnes comme variable contenant les libellés des colonnes des données et la variable 'Type client' comme variable cible d'intérêt.

Conservons les paramètres par défaut pour les règles.

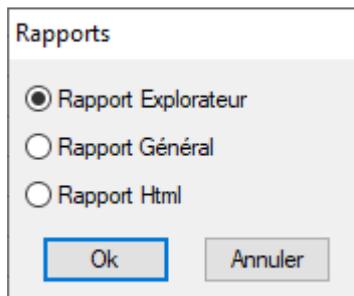
Cliquons sur le bouton Ok pour exécuter le traitement de l'analyse

Après quelques instants, la fenêtre « Rapports et Graphiques » s'affiche :

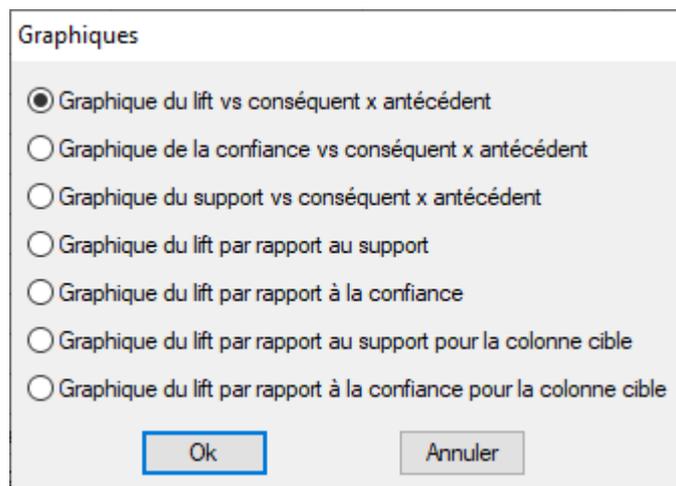


La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

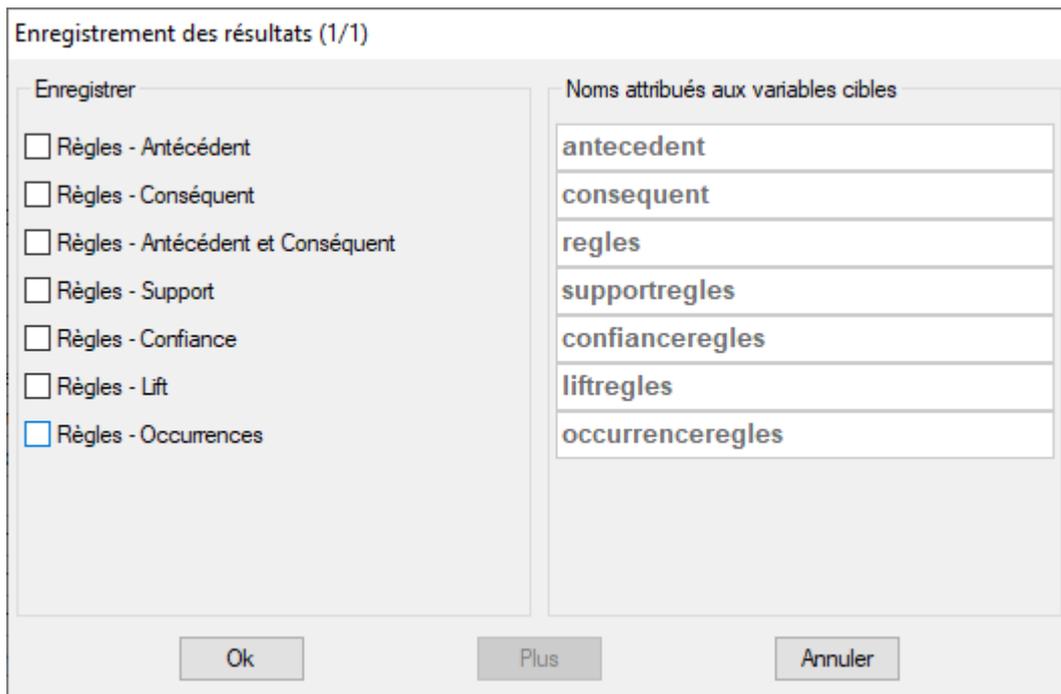
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



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



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



L'icône 'Quitter'  permet de quitter l'analyse.

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableur ou au format HTML.

L'impression des rapports fait appel à la procédure 'Aperçu avant impression'. Pour des informations sur cette procédure, voir le 'Manuel de l'Utilisateur'.

Le premier tableau indique les nombres de lignes et de colonnes du tableau des données. Il affiche la densité de la matrice lignes x colonnes (pourcentage des cellules non vides du tableau disjonctif complet formé par toutes les indicatrices des modalités des colonnes) et rappelle les paramètres définis pour la sélection des règles.

Le deuxième tableau affiche la liste des colonnes

Le troisième tableau affiche la liste des colonnes et des modalités de l'ensemble des colonnes des données.

Le quatrième tableau affiche la liste des motifs fréquents et pour chaque motif son support, son occurrence et son cardinal. Ce tableau est trié par valeurs décroissantes des supports.

Les cinquième et sixième tableaux affichent les listes des motifs fréquents fermés et maximaux.

Le tableau 'Détails des règles' affichent les règles générées triées par valeurs décroissantes du lift.

| 1 | 2 | 3 | 4 | 5 | 6 |
|----|---|---------|------------|----------|---|
| 1 | | | | | |
| 2 | MOTIFS FREQUENTS TRIES PAR SUPPORTS DECREISSANTS | | | | |
| 3 | Motif = ensemble de modalités | | | | |
| 4 | Nombre de motifs fréquents : 582 | | | | |
| 5 | Motif fréquent = motif dont le support est supérieur au support minimal choisi de 0,3 | | | | |
| 6 | Support = indicateur de fiabilité du motif | | | | |
| 7 | Occurrence = occurrence du motif | | | | |
| 8 | Cardinal = cardinal du motif | | | | |
| 9 | | | | | |
| 10 | | | | | |
| 11 | | Support | Occurrence | Cardinal | |
| 12 | {GARANTIE=AUCUNE + TRAVAILLEUR=ETRANGER} | 0,88 | 88 | 2 | |
| 13 | {RESPONSABLE=UN + TRAVAILLEUR=ETRANGER} | 0,83 | 83 | 2 | |
| 14 | {AUTRES CREDITS=AUCUN + TRAVAILLEUR=ETRANGER} | 0,76 | 76 | 2 | |
| 15 | {GARANTIE=AUCUNE + RESPONSABLE=UN } | 0,76 | 76 | 2 | |
| 16 | {GARANTIE=AUCUNE + RESPONSABLE=UN + TRAVAILLEUR=ETRANGER} | 0,75 | 75 | 3 | |
| 17 | {TYPE DE CLIENT=BON CLIENT + TRAVAILLEUR=ETRANGER} | 0,73 | 73 | 2 | |
| 18 | {LOGEMENT=PROPRIETAIRE + TRAVAILLEUR=ETRANGER} | 0,71 | 71 | 2 | |
| 19 | {GARANTIE=AUCUNE + AUTRES CREDITS=AUCUN } | 0,70 | 70 | 2 | |
| 20 | {GARANTIE=AUCUNE + AUTRES CREDITS=AUCUN + TRAVAILLEUR=ETRANGER} | 0,68 | 68 | 3 | |
| 21 | {NOMBRE CREDITS=1 CREDIT + TRAVAILLEUR=ETRANGER} | 0,67 | 67 | 2 | |
| 22 | {TYPE DE CLIENT=BON CLIENT + GARANTIE=AUCUNE } | 0,67 | 67 | 2 | |

| 1 | 2 | 3 | 4 | 5 | 6 |
|----|---|---------|---------|-----------|------------|
| 1 | | | | | |
| 2 | DETAILS DES REGLES TRIEES PAR LIFT DECREISSANTS | | | | |
| 3 | | | | | |
| 4 | Seules les règles dont le lift est supérieur ou égal à 1 sont affichées. | | | | |
| 5 | Nombre de règles : 1466 | | | | |
| 6 | La première colonne affiche Antécédent ==>> Conséquent | | | | |
| 7 | Lift = indicateur de pertinence de la règle | | | | |
| 8 | Support = indicateur de fiabilité de la règle (compris entre 0 et 1) | | | | |
| 9 | Confiance = indicateur de précision de la règle (compris entre 0 et 1) | | | | |
| 10 | Occurrence = occurrence de la règle | | | | |
| 11 | Cardinal = cardinal de la règle | | | | |
| 12 | | | | | |
| 13 | | | | | |
| 14 | | Lift | Support | Confiance | Occurrence |
| 15 | SI {LOGEMENT=PROPRIETAIRE + NOMBRE CREDITS=1 CREDIT + RESPONSABLE=UN } ALORS {HISTORIQUE CREDIT=CREDITS EN COURS } | 1,52972 | 0,35 | 0,79545 | 35 |
| 16 | SI {LOGEMENT=PROPRIETAIRE + NOMBRE CREDITS=1 CREDIT + RESPONSABLE=UN + TRAVAILLEUR=ETRANGER } ALORS {HISTORIQUE CREDIT=CREDITS EN COURS } | 1,52972 | 0,35 | 0,79545 | 35 |
| 17 | SI {NOMBRE CREDITS=1 CREDIT + EMPLOI=EMPLOYE } ALORS {HISTORIQUE CREDIT=CREDITS EN COURS } | 1,51099 | 0,33 | 0,78571 | 33 |
| 18 | SI {NOMBRE CREDITS=1 CREDIT + EMPLOI=EMPLOYE + TRAVAILLEUR=ETRANGER } ALORS {HISTORIQUE CREDIT=CREDITS EN COURS } | 1,51099 | 0,33 | 0,78571 | 33 |
| 19 | SI {GARANTIE=AUCUNE + LOGEMENT=PROPRIETAIRE + NOMBRE CREDITS=1 CREDIT + RESPONSABLE=UN } ALORS {HISTORIQUE CREDIT=CREDITS EN COURS } | 1,49038 | 0,31 | 0,77500 | 31 |
| 20 | SI {NOMBRE CREDITS=1 CREDIT + RESPONSABLE=UN } ALORS {HISTORIQUE CREDIT=CREDITS EN COURS } | 1,47665 | 0,43 | 0,76786 | 43 |
| 21 | SI {NOMBRE CREDITS=1 CREDIT + RESPONSABLE=UN + TRAVAILLEUR=ETRANGER } ALORS {HISTORIQUE CREDIT=CREDITS EN COURS } | 1,47665 | 0,43 | 0,76786 | 43 |
| 22 | SI {TYPE DE CLIENT=BON CLIENT + NOMBRE CREDITS=1 CREDIT + RESPONSABLE=UN } ALORS {HISTORIQUE CREDIT=CREDITS EN COURS } | 1,47585 | 0,33 | 0,76744 | 33 |

Le tableau 'Règles pour la colonne cible' affiche les règles qui ont pour conséquent la colonne cible d'intérêt précisée : 'Type client' dans notre exemple.

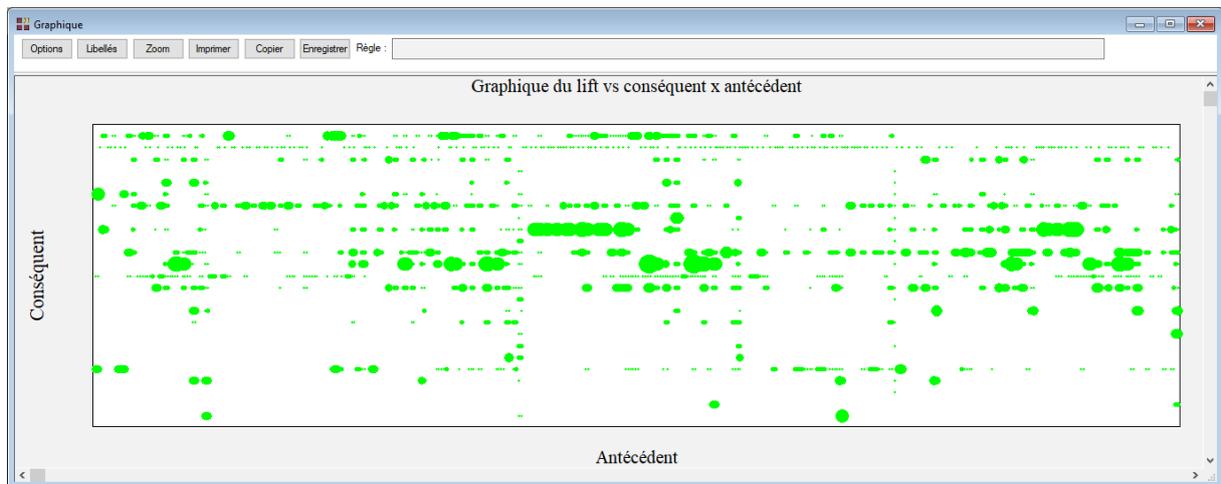
| 1 | 2 | 3 | 4 | 5 |
|----|--|---------|---------|-----------|
| 1 | | | | |
| 2 | REGLES AYANT POUR CONSEQUENT LA COLONNE CIBLE ET TRIEES PAR LIFT DECREISSANTS | | | |
| 3 | | | | |
| 4 | Colonne cible : Type de client | | | |
| 5 | Nombre de règles : 137 | | | |
| 6 | La première colonne affiche Antécédent ==>> Conséquent | | | |
| 7 | Lift = indicateur de pertinence de la règle | | | |
| 8 | Support = indicateur de fiabilité de la règle (compris entre 0 et 1) | | | |
| 9 | Confiance = indicateur de précision de la règle (compris entre 0 et 1) | | | |
| 10 | Occurrence = occurrence de la règle | | | |
| 11 | Cardinal = cardinal de la règle | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | Lift | Support | Confiance |
| 15 | SI {COMPTE CHEQUE=PAS DE COMPTE CHEQUE } ALORS {TYPE DE CLIENT=BON CLIENT } | 1,21905 | 0,32 | 0,91429 |
| 16 | SI {DUREE CREDIT=1 MOIS + RESPONSABLE=UN } ALORS {TYPE DE CLIENT=BON CLIENT } | 1,21905 | 0,32 | 0,91429 |
| 17 | SI {COMPTE CHEQUE=PAS DE COMPTE CHEQUE + TRAVAILLEUR=ETRANGER } ALORS {TYPE DE CLIENT=BON CLIENT } | 1,21569 | 0,31 | 0,91176 |
| 18 | SI {DUREE CREDIT=1 MOIS + AUTRES CREDITS=AUCUN } ALORS {TYPE DE CLIENT=BON CLIENT } | 1,21569 | 0,31 | 0,91176 |
| 19 | SI {DUREE CREDIT=1 MOIS + RESPONSABLE=UN + TRAVAILLEUR=ETRANGER } ALORS {TYPE DE CLIENT=BON CLIENT } | 1,21569 | 0,31 | 0,91176 |
| 20 | SI {DUREE CREDIT=1 MOIS } ALORS {TYPE DE CLIENT=BON CLIENT } | 1,20325 | 0,37 | 0,90244 |
| 21 | SI {DUREE CREDIT=1 MOIS + TRAVAILLEUR=ETRANGER } ALORS {TYPE DE CLIENT=BON CLIENT } | 1,19658 | 0,35 | 0,89744 |
| 22 | SI {DUREE CREDIT=1 MOIS + GARANTIE=AUCUNE } ALORS {TYPE DE CLIENT=BON CLIENT } | 1,18919 | 0,33 | 0,89189 |

Le dernier tableau affiche les occurrences des modalités de la colonne cible. Dans notre exemple, seule la modalité 'Bon client' est représentée.

| | 1 | 2 | 3 | 4 | 5 |
|---|--|-------------|------------|----------------------|---------------------|
| 1 | | | | | |
| 2 | OCCURRENCES ET FREQUENCES DES MODALITES DE LA COLONNE CIBLE | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | Occurrences | Fréquences | Occurrences cumulées | Fréquences cumulées |
| 6 | {TYPE DE CLIENT=BON CLIENT } | 137 | 100 | 137 | 100 |

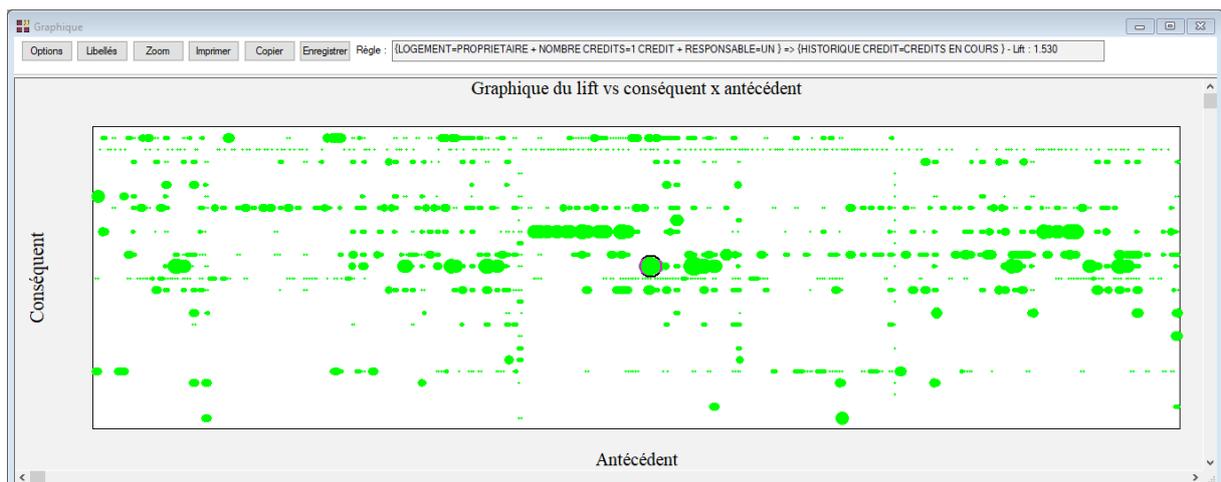
L'option Graphiques

- Graphique du lift vs conséquent x antécédent

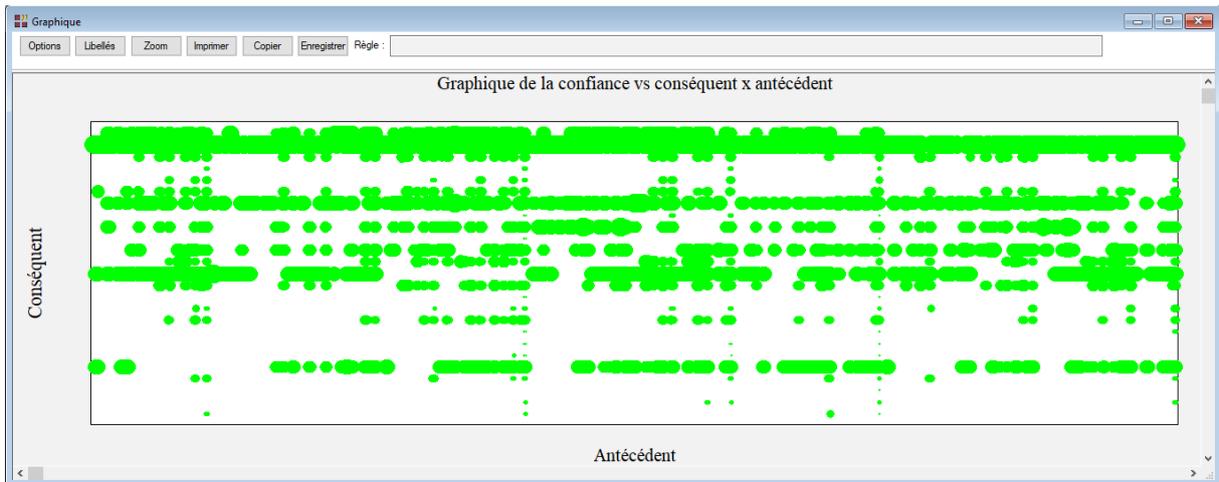


Les tailles des symboles sont proportionnelles aux valeurs du lift.

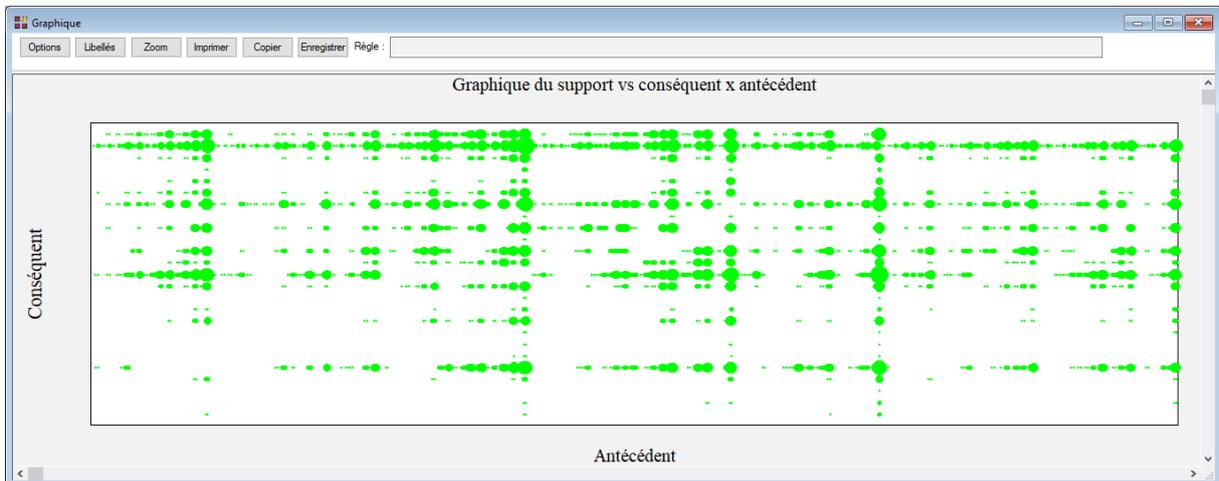
Cliquer sur un point affiche la règle et la valeur du lift dans la barre d'outils :



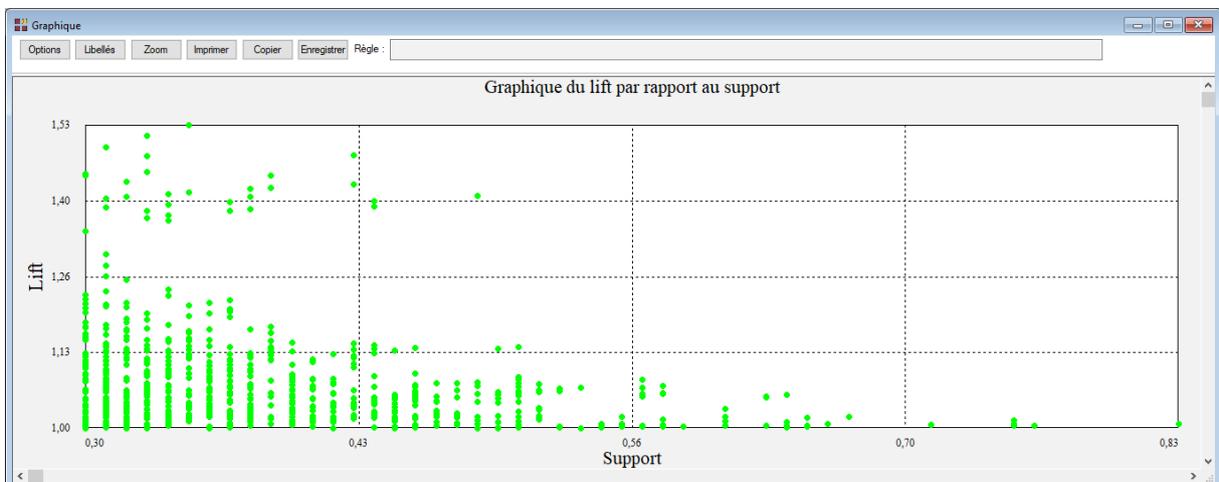
- Graphique de la confiance vs conséquent x antécédent



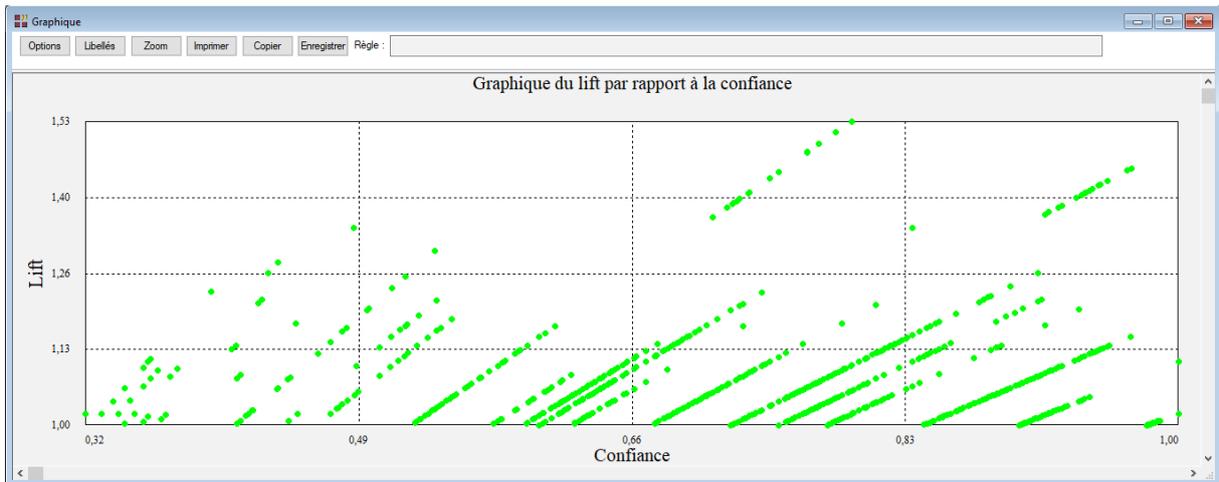
- Graphique du support vs conséquent x antécédent



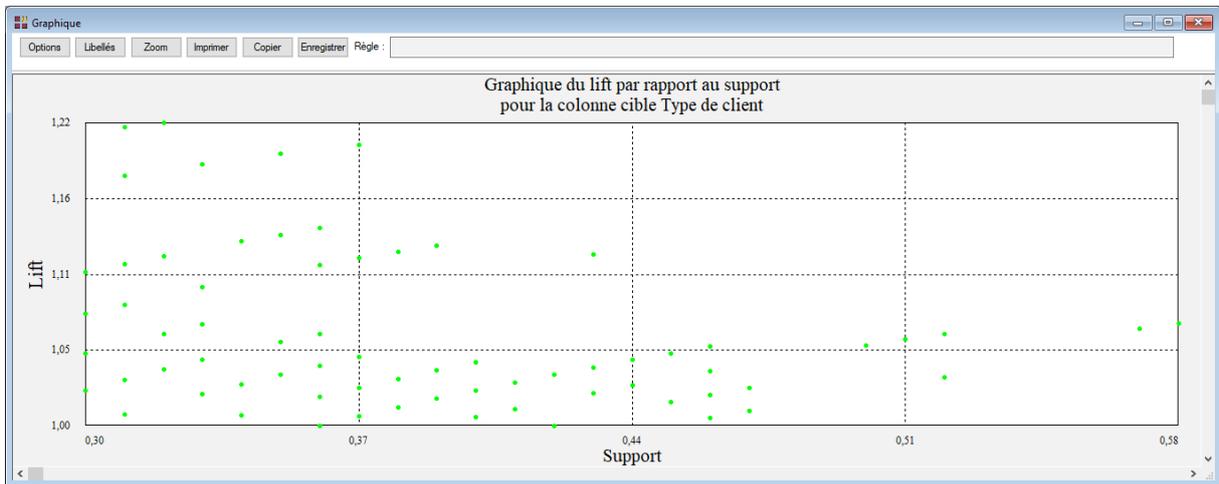
- Graphique du lift par rapport au support



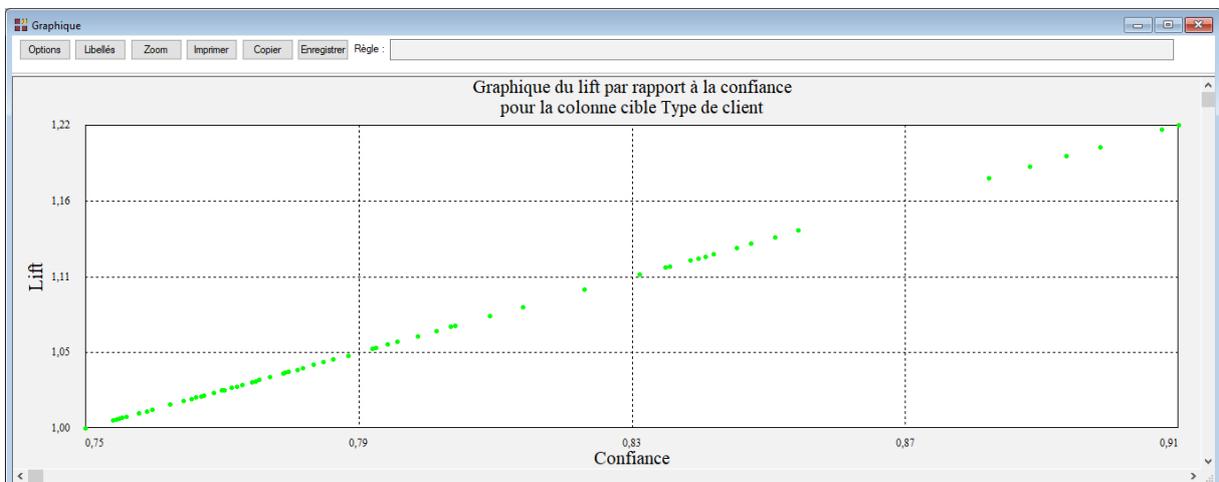
- Graphique du lift par rapport à la confiance



- Graphique du lift par rapport au support pour la colonne cible



- Graphique du lift par rapport à la confiance pour la colonne cible



Exemple 2 : Fichier Income

Le fichier Income provient du livre « The Elements of Statistical Learning ». Il contient les résultats d'une enquête concernant 6876 individus décrits par 14 caractéristiques démographiques :

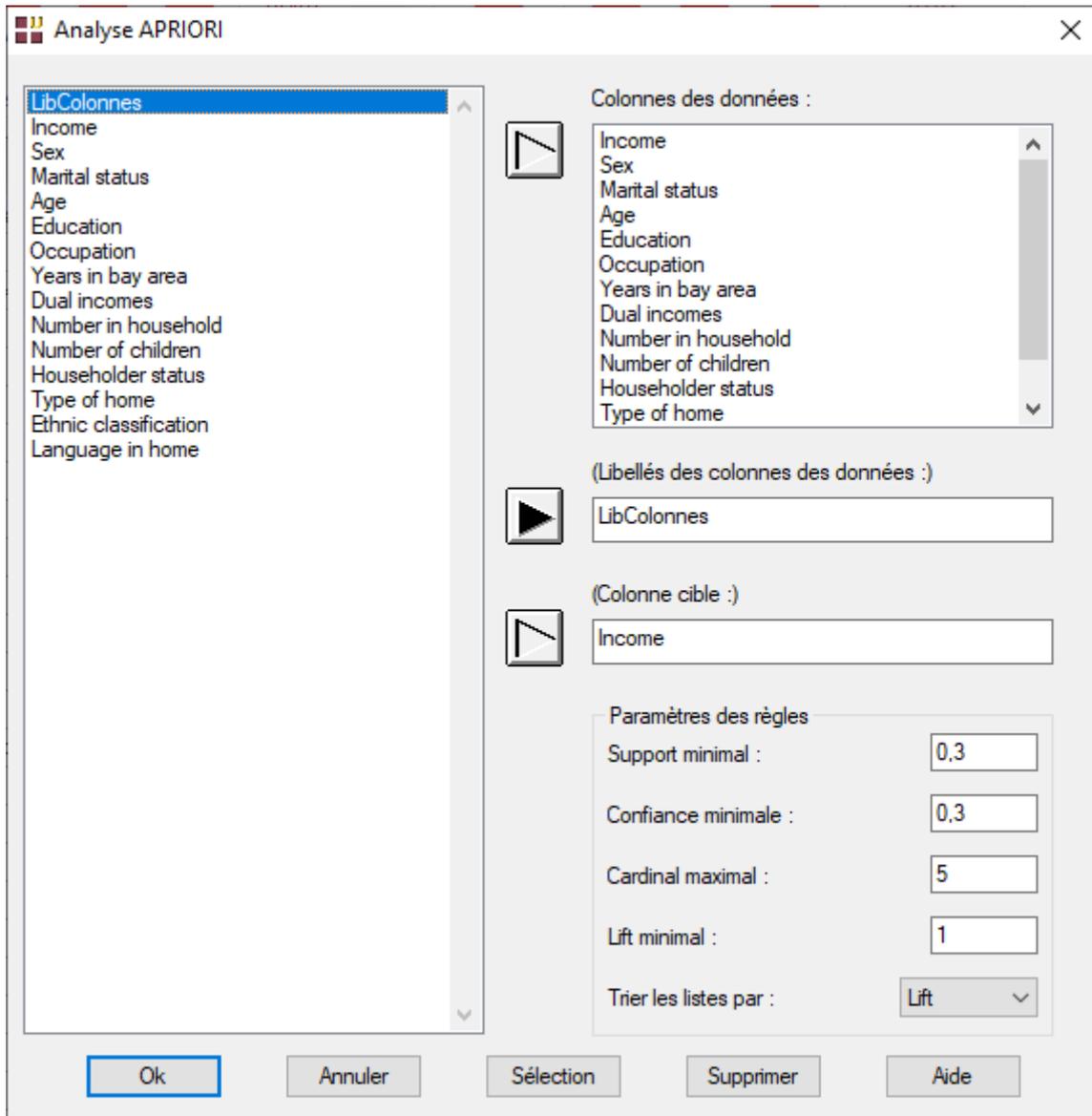
| | |
|-----------------------|--|
| Income | \$0-40000, \$40000+ |
| Sex | Female, Male |
| Marital status | Cohabitation, Divorced, Married, Single, Widowed |
| Age | 14-34, 35+ |
| Education | College graduate, No college graduate |
| Occupation | Clerical/Service, Homemaker, Laborer, Military, Professional/Managerial, Retired, Sales, Student, Unemployed |
| Years in bay area | 1-9, 10+ |
| Dual incomes | No, Not married, Yes |
| Number in household | 1, 2+ |
| Number of children | 0, 1+ |
| Householder status | Live with parents/family, Own, Rent |
| Type of home | Apartment, Condominium, House, Mobile house, Other |
| Ethnic classification | American indian, Asian, Black, East indian, Hispanic, Other, Pacific islander, White |
| Language in home | English, Other, Spanish |

Renseignons la boîte de dialogue comme montré ci-après (on sélectionne les colonnes de données de 'Income' à 'Langage in home').

Conservons les paramètres par défaut pour les règles.

Cliquons sur le bouton Ok pour exécuter le traitement de l'analyse.

Après quelques instants, la fenêtre « Rapports et Graphiques » s'affiche.



Rapports et Graphiques

Rapport APRIORI

- Libellés des colonnes
- Libellés des modalités
- Motifs fréquents
- Motifs fréquents fermés
- Motifs fréquents maximaux
- Détails des règles
- Règles pour la colonne cible
- Occurrences modalités cible

| | | | | | | | | | |
|----|--|--|--|--|--|--|--|--|--|
| 1 | | | | | | | | | |
| 2 | (C) UNWIN version 9.3.5 | | | | | | | | |
| 3 | | | | | | | | | |
| 4 | DATE : 11/03/2022 | | | | | | | | |
| 5 | ORDINATEUR : LAPTOP-8CVDBRT1 | | | | | | | | |
| 6 | UTILISATEUR : Christian | | | | | | | | |
| 7 | FICHIER(S) DE DONNEES OUVERT(S) : INCOME.SGD | | | | | | | | |
| 8 | | | | | | | | | |
| 9 | RESULTATS DE L'ANALYSE APRIORI | | | | | | | | |
| 10 | | | | | | | | | |
| 11 | Sélection : | | | | | | | | |
| 12 | Aucune | | | | | | | | |
| 13 | | | | | | | | | |
| 14 | Nombre de lignes de données : 6876 | | | | | | | | |
| 15 | Nombre de colonnes de données : 14 | | | | | | | | |
| 16 | Nombre de modalités des colonnes : 50 | | | | | | | | |
| 17 | Densité de la matrice lignes x colonnes : 0,280 | | | | | | | | |
| 18 | Variable cible : INCOME | | | | | | | | |
| 19 | Support minimal pour la construction des règles : 0,3 | | | | | | | | |
| 20 | Confiance minimale pour la construction des règles : 0,3 | | | | | | | | |
| 21 | Cardinal maximal pour la construction des règles : 5 | | | | | | | | |
| 22 | Lift minimal pour l'affichage des règles : 1 | | | | | | | | |

< > Rapport Explorateur | <

Règles triées par valeurs décroissantes du lift

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----|---|---------|---------|-----------|------------|----------|
| 1 | DETAILS DES REGLES TRIEES PAR LIFT DECREISSANTS | | | | | |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | Seules les règles dont le lift est supérieur ou égal à 1 sont affichées. | | | | | |
| 5 | Nombre de règles : 281 | | | | | |
| 6 | La première colonne affiche Antécédent ==>> Conséquent | | | | | |
| 7 | Lift = indicateur de pertinence de la règle | | | | | |
| 8 | Support = indicateur de fiabilité de la règle (compris entre 0 et 1) | | | | | |
| 9 | Confiance = indicateur de précision de la règle (compris entre 0 et 1) | | | | | |
| 10 | Occurrence = occurrence de la règle | | | | | |
| 11 | Cardinal = cardinal de la règle | | | | | |
| 12 | | | | | | |
| 13 | | | | | | |
| 14 | | Lift | Support | Confiance | Occurrence | Cardinal |
| 15 | SI {INCOME=\$0-\$40,000 + AGE=14-34 + DUAL INCOMES=NOT MARRIED} ALORS {MARITAL STATUS=SINGLE } | 2,04906 | 0,30003 | 0,83828 | 2063 | 4 |
| 16 | SI {AGE=14-34 + DUAL INCOMES=NOT MARRIED} ALORS {MARITAL STATUS=SINGLE } | 2,03198 | 0,36547 | 0,83129 | 2513 | 3 |
| 17 | SI {AGE=14-34 + DUAL INCOMES=NOT MARRIED + LANGUAGE IN HOME=ENGLISH} ALORS {MARITAL STATUS=SINGLE } | 2,01407 | 0,32403 | 0,82396 | 2228 | 4 |
| 18 | SI {INCOME=\$0-\$40,000 + DUAL INCOMES=NOT MARRIED} ALORS {MARITAL STATUS=SINGLE } | 1,74674 | 0,33028 | 0,71460 | 2271 | 3 |
| 19 | SI {EDUCATION=NO COLLEGE GRADUATE + DUAL INCOMES=NOT MARRIED} ALORS {MARITAL STATUS=SINGLE } | 1,72862 | 0,32068 | 0,70718 | 2205 | 3 |
| 20 | SI {INCOME=\$0-\$40,000 + AGE=14-34} ALORS {MARITAL STATUS=SINGLE } | 1,67979 | 0,30003 | 0,68721 | 2063 | 3 |
| 21 | SI {MARITAL STATUS=SINGLE } ALORS {DUAL INCOMES=NOT MARRIED} | 1,67137 | 0,40910 | 1,00000 | 2813 | 2 |
| 22 | SI {DUAL INCOMES=NOT MARRIED} ALORS {MARITAL STATUS=SINGLE } | 1,67137 | 0,40910 | 0,68376 | 2813 | 2 |

Règles ayant pour conséquent la colonne cible

| 1 | 2 | 3 | 4 | 5 | 6 |
|----|--|---------|---------|-----------|----------|
| 1 | REGLES AYANT POUR CONSEQUENT LA COLONNE CIBLE ET TRIEES PAR LIFT DECREISSANTS | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | Colonne cible : Income | | | | |
| 5 | Nombre de règles : 24 | | | | |
| 6 | La première colonne affiche Antécédent ==>> Conséquent | | | | |
| 7 | Lift = indicateur de pertinence de la règle | | | | |
| 8 | Support = indicateur de fiabilité de la règle (compris entre 0 et 1) | | | | |
| 9 | Confiance = indicateur de précision de la règle (compris entre 0 et 1) | | | | |
| 10 | Occurrence = occurrence de la règle | | | | |
| 11 | Cardinal = cardinal de la règle | | | | |
| 12 | | | | | |
| 13 | | | | | |
| 14 | | Lift | Support | Confiance | Cardinal |
| 15 | SI {EDUCATION=NO COLLEGE GRADUATE + DUAL INCOMES=NOT MARRIED} ALORS {INCOME=\$0-\$40,000} | 1,32006 | 0,37260 | 0,82168 | 2562 |
| 16 | SI {MARITAL STATUS=SINGLE + AGE=14-34} ALORS {INCOME=\$0-\$40,000} | 1,31886 | 0,30003 | 0,82093 | 2063 |
| 17 | SI {MARITAL STATUS=SINGLE + AGE=14-34 + DUAL INCOMES=NOT MARRIED} ALORS {INCOME=\$0-\$40,000} | 1,31886 | 0,30003 | 0,82093 | 2063 |
| 18 | SI {EDUCATION=NO COLLEGE GRADUATE + DUAL INCOMES=NOT MARRIED + LANGUAGE IN HOME=ENGLISH} ALORS {INCOME=\$0-\$40,000} | 1,31816 | 0,33304 | 0,82049 | 2290 |
| 19 | SI {AGE=14-34 + DUAL INCOMES=NOT MARRIED} ALORS {INCOME=\$0-\$40,000} | 1,30787 | 0,35791 | 0,81409 | 2461 |
| 20 | SI {AGE=14-34 + DUAL INCOMES=NOT MARRIED + LANGUAGE IN HOME=ENGLISH} ALORS {INCOME=\$0-\$40,000} | 1,30769 | 0,32010 | 0,81398 | 2201 |
| 21 | SI {MARITAL STATUS=SINGLE } ALORS {INCOME=\$0-\$40,000} | 1,29700 | 0,33028 | 0,80732 | 2271 |
| 22 | SI {MARITAL STATUS=SINGLE + DUAL INCOMES=NOT MARRIED} ALORS {INCOME=\$0-\$40,000} | 1,29700 | 0,33028 | 0,80732 | 2271 |

Occurrences des modalités de la variable cible

| 1 | 2 | 3 | 4 | 5 |
|---|---|-------------|------------|----------------------|
| 1 | OCCURRENCES ET FREQUENCES DES MODALITES DE LA COLONNE CIBLE | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | Occurrences | Fréquences | Occurrences cumulées |
| 6 | {INCOME=\$0-\$40,000} | 23 | 95,83333 | 23 |
| 7 | {INCOME=\$40,000+ } | 1 | 4,16667 | 24 |

Exemple 3 : Fichier Ventess2

Une enquête a été effectuée concernant 1596 clientes d'une société de vente par correspondance. Les clientes ont été interrogées sur les thèmes suivants :

| <u>Thème</u> | <u>Libellé</u> |
|--|--------------------|
| 1. Satisfaction des clientes | Satisfaction |
| 2. Commander d'autres produits | Autres commandes |
| 3. Conseiller les produits à des amies | Conseils amies |
| 4. Gestion des retours/échanges | Echanges |
| 5. Exactitude des factures | Factures |
| 6. Conformité du contenu du colis | Colis |
| 7. Rapidité de la livraison | Rapidité livraison |
| 8. Qualité de la livraison | Qualité livraison |
| 9. Diversité des modes de règlement | Règlement |
| 10. Flexibilité en cas d'annulation d'une commande | Annulation |
| 11. Facilité à passer une commande | Commande |
| 12. Montant des frais de port et d'emballage | Port |
| 13. Conformité entre photo et réalité des produits | Conformité |
| 14. Attrait, qualité et diversité des cadeaux | Cadeaux |
| 15. Possibilité d'essayer les produits | Echantillons |
| 16. Prix des produits | Prix |
| 17. Diversité et renouvellement des produits | Choix |
| 18. Qualité et efficacité des produits | Qualité produit |
| 19. Attrait et caractère innovant des produits | Attrait produit |
| 20. Attrait et clarté des offres promotionnelles | Promotions |
| 21. Relation téléphonique pour renseignements | Conseil téléphone |
| 22. Relation téléphonique pour commande | Commande téléphone |
| 23. Prise de contact en cas de problèmes | Prise contact |
| 24. Age des clientes | Age |

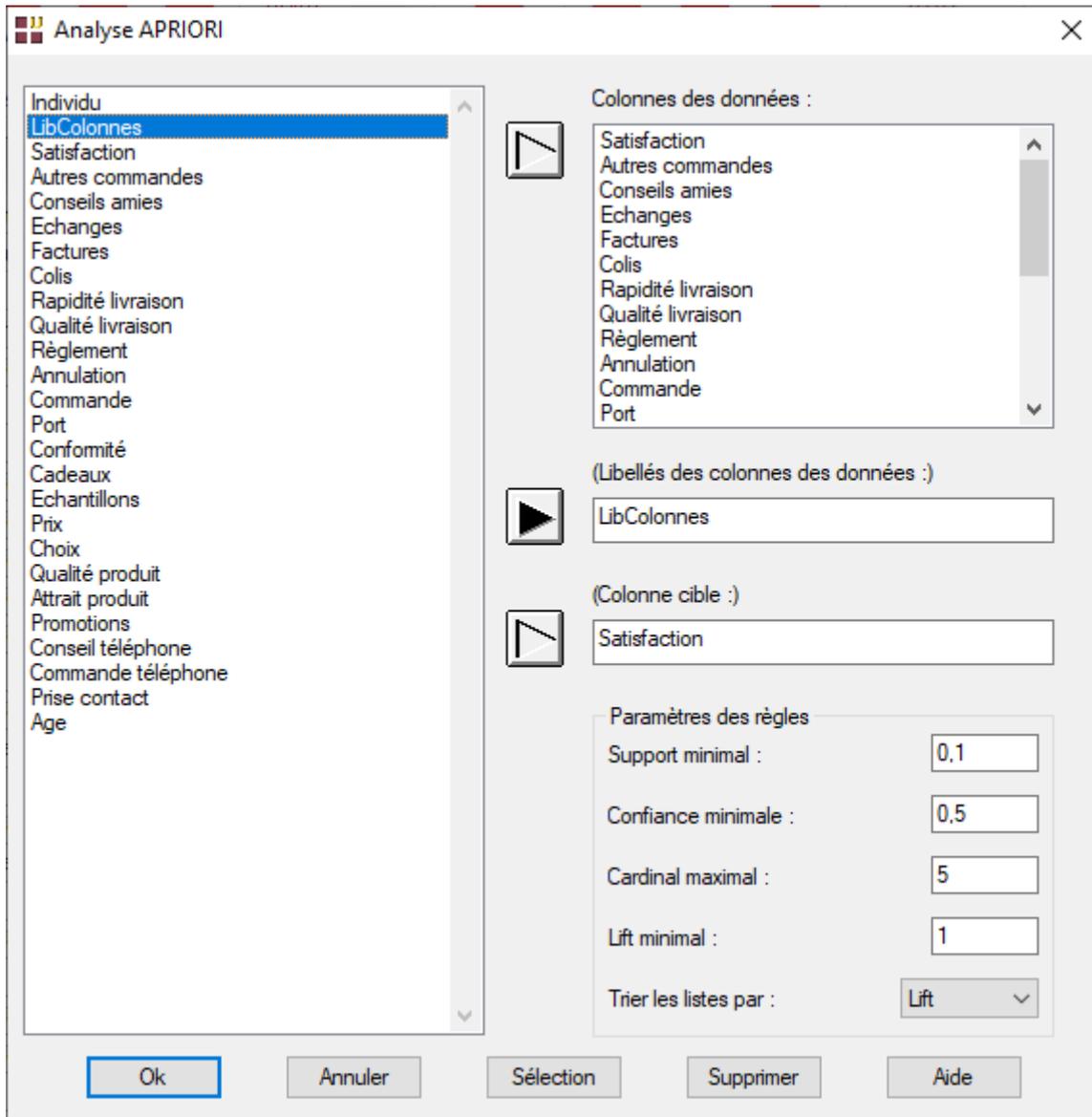
Chaque thème est codé 'Très satisfaite', 'Satisfaite', 'Indifférente', 'Mécontente', 'Très mécontente'.

L'âge des clientes est codé de la façon suivante : plus de 50 ans, 30 à 50 ans, moins de 30 ans

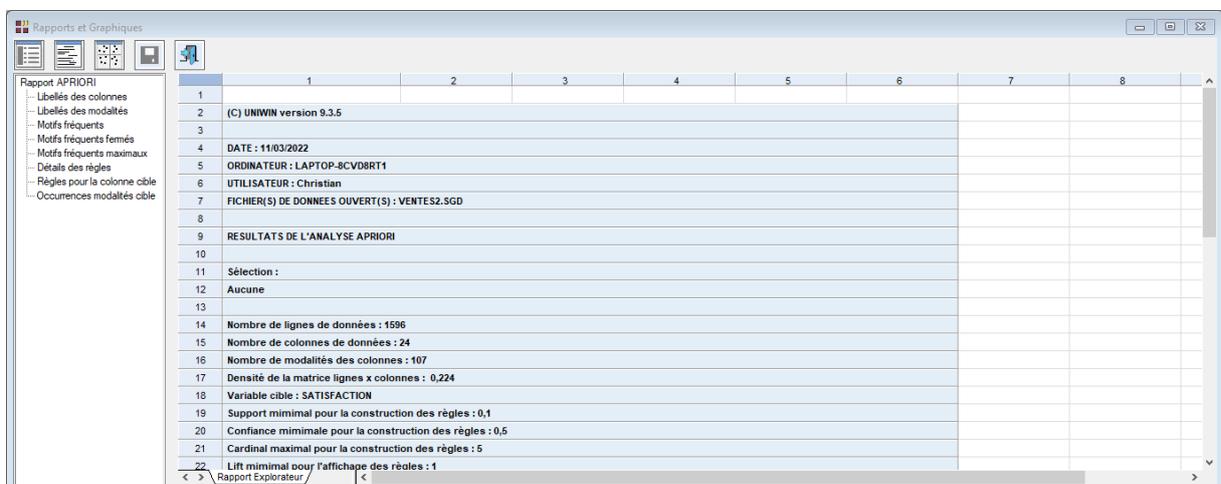
Renseignons la boîte de dialogue comme montré ci-après.

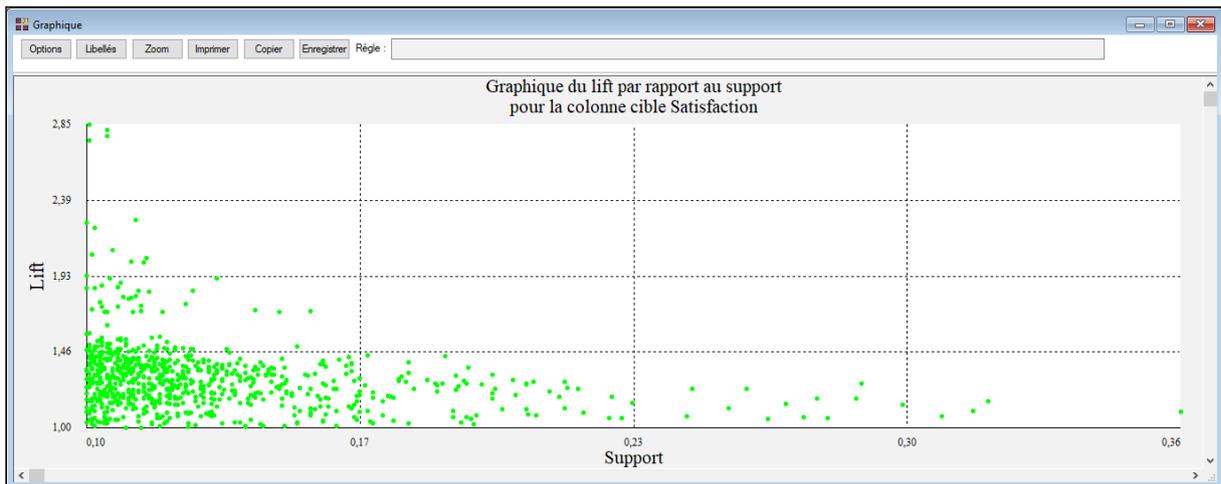
Sélectionnons les colonnes des données de 'Satisfaction' à 'Age', la variable 'LibColonnes' comme variable contenant les libellés des colonnes des données, la variable 'Satisfaction' comme colonne cible et définissons le support minimal à 0,1 et la confiance minimale à 0,5.

Cliquons sur le bouton Ok pour exécuter le traitement de l'analyse.



Après l'attente due aux nombreux calculs, la fenêtre « Rapports et Graphiques » s'affiche.





Les variables créées par la procédure

Voici la liste des variables créées par la procédure.

| <i>Variable</i> | <i>Contenu</i> |
|------------------|------------------------|
| antecedent | antécédents des règles |
| consequent | conséquents des règles |
| regles | règles complètes |
| supportregles | supports des règles |
| confianceregles | confiances des règles |
| liftregles | lifts des règles |
| occurrenceregles | occurrences des règles |

Références

[Documentation du package R 'arules' \(2021\)](#)

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

Exemple 1 : Crédit3

Le fichier Crédit d'origine est disponible sur le site « UCI Machine Learning Repository ». Il a été conçu par le Professeur Dr. Hans Hofmann de l'Institut für Statistik und Okonometrie, Universität Hamburg

Exemple 2 : Income

Hastie, T., Tibshirani, R. et Friedman, J. (2001) The Elements of Statistical Learning. Springer-Verlag.