

IBM Analytical Decision Management
Version 18 Edition 0

IBM SPSS Modeler Advantage

IBM

Important

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 99.

Informations sur le produit

LE PRESENT DOCUMENT EST LIVRE EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE.

Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. Les informations qui y sont fournies sont susceptibles d'être modifiées avant que les produits décrits ne deviennent eux-mêmes disponibles. En outre, il peut contenir des informations ou des références concernant certains produits, logiciels ou services non annoncés dans ce pays. Cela ne signifie cependant pas qu'ils y seront annoncés.

Pour plus de détails, pour toute demande d'ordre technique, ou pour obtenir des exemplaires de documents IBM, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial.

Vous pouvez également consulter les serveurs Internet suivants :

- <http://www.fr.ibm.com> (serveur IBM en France)
- <http://www.ibm.com/ca/fr> (serveur IBM au Canada)
- <http://www.ibm.com> (serveur IBM aux Etats-Unis)

*Compagnie IBM France
Direction Qualité
17, avenue de l'Europe
92275 Bois-Colombes Cedex*

© Copyright IBM France 2015. Tous droits réservés.

Cette édition s'applique à la version 18.0.0 d'IBM Analytical Decision Management et à toutes les éditions et modifications ultérieures, sauf mention contraire dans les éditions suivantes.

© Copyright IBM Corporation 2010, 2015.

Table des matières

Avis aux lecteurs canadiens	v	Édition des valeurs de champ de données	47
Préface	vii	Mappage de champs	47
Chapitre 1. A propos d'IBM Analytical Decision Management	1	Calcul des champs	48
Présentation du produit	1	Générateur de formules	49
Nouveautés de la version 18	1	Sélection de fonctions	50
Versions précédentes.	2	Sélection de champs et de valeurs	50
Chapitre 2. Exemple de modélisation des réponses des clients	11	Vérification des expressions	51
IBM SPSS Modeler Advantage	11	Utilisation des expressions qui renvoient un booléen.	51
Optimisation du profit pour les campagnes de publipostage	11	Chapitre 5. Utilisation des règles	53
Exemples de fichiers	11	Création des règles	53
Lancement de l'application	13	Définition de règles de segment	53
Configuration des sources de données	13	Définition de règles de sélection	55
Génération du modèle.	15	Exportation et réutilisation de règles	56
Evaluation du modèle	17	Ajout d'annotations.	57
Modélisation de règle prédictive	20	Utilisation de règles IBM Operational Decision Management	57
Test du modèle	21	Chapitre 6. Génération de modèles prédictifs.	59
Evaluation du modèle	21	A propos des modèles prédictifs	59
Exemple de modélisation de groupement	22	Génération de modèles	59
Exemple de modélisation d'association	23	Génération d'un modèle prédictif	59
Chapitre 3. Navigation dans l'espace de travail	27	Résultats de la modélisation prédictive	60
Changement de la langue affichée	27	Génération d'un modèle de règle prédictif	61
Lancement d'application	27	Génération d'un modèle de groupement.	63
Page d'accueil	28	Génération d'un modèle d'association	64
Galerie	28	Paramètres facultatifs de modèle	65
Définition des préférences	29	Evaluation de modèles	69
Ouverture et enregistrement de projets, de modèles et de règles	31	Simulation du profit maximal	71
Ouverture depuis et enregistrement dans le référentiel	31	Test du modèle	72
Utilisation des libellés	32	Utilisation de scores de modèle dans des applications	72
Verrouillage des projets	33	Chapitre 7. Evaluation de modèles.	75
Propriétés	33	Evaluation (scoring) de modèles vers une table de base de données, un fichier, un serveur analytique ou un serveur IBM Cognos BI	75
Chapitre 4. Gestion des sources de données	35	Sélection de données et de sous-ensembles	77
Création d'une source de données	37	Sélection des enregistrements à scorer	77
Source Fichier	38	Sélection de la destination de l'évaluation (scoring)	77
Source Base de données	41	Sélection de champs de sortie	81
Source Vue de données	42	Évaluation de modèles	81
Source IBM Cognos BI.	42	Rapport récapitulatif d'évaluation	82
Source Analytic Server.	44	Chapitre 8. Administration des applications	83
Sélection de champs d'entrée	45	Options générales	84
Aperçu des données	45	Contrôle des raccourcis d'application	85
Présentation des données.	45	Masquage de la page d'accueil	85
Niveaux de mesure.	46	Verrouillage des options de données	85
		Verrouillage des options de destination d'évaluation	86

Téléchargement de métadonnées de projet à utiliser pour créer des règles externes	86	Dispositions de la documentation du produit. . .	101
Edition de la définition de schéma XML.	88	Index	103
Chapitre 9. Obtention de l'aide	93		
Documentation disponible	93		
Annexe A. Partage des flux	95		
Annexe B. Accessibilité	97		
Remarques	99		
Marques	100		

Avis aux lecteurs canadiens

Le présent document a été traduit en France. Voici les principales différences et particularités dont vous devez tenir compte.

Illustrations

Les illustrations sont fournies à titre d'exemple. Certaines peuvent contenir des données propres à la France.

Terminologie

La terminologie des titres IBM peut différer d'un pays à l'autre. Reportez-vous au tableau ci-dessous, au besoin.

IBM France	IBM Canada
ingénieur commercial	représentant
agence commerciale	succursale
ingénieur technico-commercial	informaticien
inspecteur	technicien du matériel

Claviers

Les lettres sont disposées différemment : le clavier français est de type AZERTY, et le clavier français-canadien de type QWERTY.

OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

- les pages de codes 850 (multilingue) et 863 (français-canadien),
- le code pays 002,
- le code clavier CF.

Nomenclature

Les touches présentées dans le tableau d'équivalence suivant sont libellées différemment selon qu'il s'agit du clavier de la France, du clavier du Canada ou du clavier des États-Unis. Reportez-vous à ce tableau pour faire correspondre les touches françaises figurant dans le présent document aux touches de votre clavier.

France	Canada	Etats-Unis
⌫ (Pos1)	⌫	Home
Fin	Fin	End
⬆ (PgAr)	⬆	PgUp
⬇ (PgAv)	⬇	PgDn
Inser	Inser	Ins
Suppr	Suppr	Del
Echap	Echap	Esc
Attn	Intrp	Break
Impr écran	ImpEc	PrtSc
Verr num	Num	Num Lock
Arrêt défil	Défil	Scroll Lock
🔒 (Verr maj)	FixMaj	Caps Lock
AltGr	AltCar	Alt (à droite)

Brevets

Il est possible qu'IBM détienne des brevets ou qu'elle ait déposé des demandes de brevets portant sur certains sujets abordés dans ce document. Le fait qu'IBM vous fournisse le présent document ne signifie pas qu'elle vous accorde un permis d'utilisation de ces brevets. Vous pouvez envoyer, par écrit, vos demandes de renseignements relatives aux permis d'utilisation au directeur général des relations commerciales d'IBM, 3600 Steeles Avenue East, Markham, Ontario, L3R 9Z7.

Assistance téléphonique

Si vous avez besoin d'assistance ou si vous voulez commander du matériel, des logiciels et des publications IBM, contactez IBM direct au 1 800 465-1234.

Préface

A propos d'IBM Business Analytics

Le logiciel IBM Business Analytics fournit des informations complètes, cohérentes et précises à l'attention des décideurs cherchant à améliorer leur performance métier. Un portefeuille complet de solutions de business intelligence, d'analyses prédictives, de performance financière et de gestion de la stratégie, et d'applications analytiques permet une connaissance claire et immédiate et offre des possibilités d'actions sur les performances actuelles et la capacité à prédire les résultats futurs. Appuyées par des solutions riches pour l'industrie, des pratiques éprouvées et des services professionnels, les entreprises peuvent obtenir une productivité optimale, des décisions automatisées fiables et ainsi produire de meilleurs résultats, quelle que soit leur taille.

Le logiciel IBM SPSS Predictive Analytics fait partie intégrante de ce portefeuille et permet aux entreprises de prédire les événements futurs et d'agir en conséquence de façon proactive dans le but de produire de meilleurs résultats. Des clients du monde entier, qu'ils soient du secteur commercial, gouvernemental ou de l'éducation, se fient à la technologie IBM® SPSS comme un avantage compétitif pour attirer, retenir et augmenter le nombre de clients tout en réduisant la fraude et les risques. En utilisant le logiciel IBM SPSS dans leurs opérations quotidiennes, les organisations deviennent des entreprises prédictives, capables de diriger et d'automatiser les décisions pour répondre aux objectifs commerciaux et obtenir un avantage concurrentiel mesurable. Pour plus d'informations ou pour contacter un interlocuteur IBM, visitez le site <http://www.ibm.com/spss>.

Support technique

L'assistance technique est disponible pour les clients du service de maintenance. Les clients peuvent contacter l'assistance technique pour obtenir de l'aide concernant l'utilisation des produits IBM Corp. ou l'installation dans l'un des environnements matériels pris en charge. Pour contacter le support technique, visitez le site IBM Corp. à l'adresse <http://www.ibm.com/support>. Votre nom, celui de votre société, ainsi que votre contrat de prise en charge vous seront demandés.

Chapitre 1. A propos d'IBM Analytical Decision Management

Présentation du produit

IBM Analytical Decision Management apporte l'avantage de l'analyse prédictive aux problèmes métiers réels, vous permettant de générer des applications personnalisées adaptées à vos clients ou à l'industrie. Bien que les applications soient généralement configurées pour résoudre des problèmes très spécifiques, elles sont toutes basées sur un ensemble commun de capacités :

- Automatiser des décisions avec des règles métier.
- Ajouter des connaissances à l'aide de modèles prédictifs.
- Utiliser la définition des priorités, l'optimisation ou la simulation pour aboutir à la meilleure décision en fonction des éléments précédents.

Un certain nombre d'applications intégrées sont disponibles, adaptées à la résolution de problèmes métier spécifiques. Pour plus d'informations, contactez votre représentant commercial.

Nouveautés de la version 18

IBM Analytical Decision Management 18 (18.0.0.0) a ajouté les fonctions suivantes :

Une description des nouvelles fonctions principales est fournie ci-dessous :

Exécution de scénario. Une nouvelle icône de propriétés dans la boîte de dialogue Gérer les scénarios vous permet de consulter les détails relatifs à un scénario, tels que l'heure de début, l'heure de fin et le temps d'exécution.

Journal CPLEX. Si vous souhaitez enregistrer le journal CPLEX dans un fichier du système de fichiers d'IBM SPSS Modeler Server à des fins d'identification et résolution des problèmes, une nouvelle option **Réacheminer le fichier journal du moteur CPLEX vers le fichier** est disponible dans la boîte de dialogue Options avancées, pour les applications qui disposent d'un onglet Optimiser. Veuillez noter que dans les rares cas où une application CPLEX est déployée vers le Service d'évaluation, il vous faut désactiver la journalisation pour améliorer les performances.

Chargement en vrac et taille de validation du lot. Dans l'onglet Déployer, lors de l'envoi vers une base de données, vous pouvez désormais définir la taille de lot et activer le chargement en vrac. Cela permet d'améliorer les performances lors de l'évaluation d'une table conséquente et de l'écriture des résultats dans une base de données.

Présentation des données en contexte. La Présentation des données s'affiche désormais dans le cas où la zone est en cours d'utilisation au sein d'un projet. Par exemple, les règles précédemment créées dans le projet sont prises en compte.

Nouvelle contrainte d'optimisation. Une nouvelle contrainte d'optimisation a été ajoutée à l'application IBM Analytical Decision Management for Campaign Optimization pour *Nombre d'offres minimum devant être acceptées pour ce type d'offre*.

Améliorations de la mémoire. Des améliorations quant à la mémoire ont été apportées à l'application IBM Analytical Decision Management for Campaign Optimization. Pour plus d'informations, voir **Meilleures pratiques en matière de performances > Optimisation CPLEX > Mémoire minimale requise** dans le *Guide de performances et d'identification et de résolution des problèmes*.

Versions précédentes

Version 17

IBM Analytical Decision Management 17 (17.0.0.0) a ajouté les fonctions suivantes :

Une description des principales nouvelles fonctions ajoutées dans la version 17 est fournie ci-dessous :

Générateur de formules. Le Générateur de formules a été amélioré pour autoriser des expressions plus complexes et plus puissantes. Le Générateur de formules dans IBM Analytical Decision Management assure désormais mieux la mise en correspondance du Générateur de formules dans le client IBM SPSS Modeler. Voir http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/adm_user_guide_ddita/configurableapps/dm_expressions.dita.

Onglet Dérivé. Dans la plupart des applications IBM Analytical Decision Management, vous pouvez ajouter des champs au modèle de données du projet, tels que des formules, des sorties de modèle, des champs créés à partir de règles de segment ou des champs de tables secondaires. Le nouveau sous-onglet **Dérivé** de l'onglet Données affiche tous les champs à inclure en tant qu'élément des données. Voir http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/adm_user_guide_ddita/configurableapps/dm_app_derived.dita.

Section Tables secondaires. Une nouvelle section Tables secondaires a été ajoutée à la plupart des applications IBM Analytical Decision Management. Elle remplace les différents liens **Ajouter/modifier des tables supplémentaires** des éditions précédentes. Voir http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/adm_user_guide_ddita/configurableapps/dm_app_additional_fields.dita.

Authentification multi-facteur. Une fonctionnalité prenant en charge l'authentification multi-facteur (MFA) a été ajoutée à IBM SPSS Collaboration and Deployment Services 7, IBM SPSS Modeler 17 et IBM Analytical Decision Management 17. Ces améliorations peuvent améliorer la sécurité pour les environnements d'entreprise à grande échelle. Par exemple, des produits SPSS peuvent être intégrés à une infrastructure de connexion Kerberos unique existante d'une société et à des infrastructures SmartCard (quand les employés d'une société utilisent des badges de sécurité et des lecteurs de cartes pour l'authentification avec la machine et les applications, par exemple). Voir http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/adm_install_guide_ddita/configurableapps/dms_install_security.dita.

Mappage de source de données. Lorsque c'est possible, IBM Analytical Decision Management mappe désormais automatiquement les types de stockage compatibles quand un utilisateur ajoute une source de données supplémentaire. Si le type n'est pas compatible, l'utilisateur sera autorisé à convertir les types.

Types de stockage de fichier. Les utilisateurs peuvent désormais définir le type et le format de stockage lors de la lecture de fichiers de données de format fixe et variable, pour s'assurer que des valeurs sont lues correctement. Par exemple, si le type de la source de données est **Fichier** et le format de fichier est **Champs de longueur variable en texte brut** lors de l'ajout d'une nouvelle source de données dans IBM Analytical Decision Management, la boîte de dialogue de source de données possède désormais une nouvelle section **Spécifier le format de données** qui permet à l'utilisateur d'indiquer le type d'entrée et le format de chaque champ. Cette procédure est similaire à la fonctionnalité disponible dans le client IBM SPSS Modeler.

Récapitulatif du projet. Pour les applications dotées d'un onglet Déployer, une nouvelle section **Récapitulatif du projet** a été ajoutée. Utilisez cette section pour valider le projet et afficher les informations récapitulatives importantes concernant le projet, y compris les erreurs détectées pouvant avoir une incidence sur le déploiement. Le récapitulatif de projet affiche des informations telles que toutes les références d'attribut et toutes les références de référentiel, et peut être exporté sous la forme d'un fichier *.csv. Vous pouvez également vous déplacer dans les autres onglets de l'application pour corriger les erreurs, puis revenir au récapitulatif de projet dans l'onglet Déployer.

Références de modèles. Quand vous référencez des modèles depuis une application IBM Analytical Decision Management, le comportement de la liaison du modèle change. Quand vous cliquez sur la liaison de référence d'un modèle, la boîte de dialogue Visualiseur de modèle s'ouvre, vous permettant d'afficher des détails au sujet du modèle. Vous pouvez désormais également cliquer avec le bouton droit de la souris sur la référence du modèle pour sélectionner l'option **Référence** ou **Version**. Quand vous cliquez sur une référence de règle, l'afficheur de règle s'ouvre (comme dans les versions précédentes). Vous pouvez également cliquer avec le bouton droit de la souris sur une référence de règle et sélectionner **Référence**.

Suffixe des données d'identification d'Analytic Server. Les administrateurs peuvent utiliser l'IBM SPSS Collaboration and Deployment Services Deployment Manager basé sur navigateur pour configurer un suffixe à ajouter aux données d'identification d'Analytic Server. Ceci dérivera le nom des données d'identification qui seront utilisées pour les connexions Analytic Server dans IBM Analytical Decision Management.

Données d'identification d'IBM Cognos. Lors de l'importation des données de base de données IBM Cognos BI ou des rapports de type Liste en vue de leur utilisation dans IBM Analytical Decision Management, vous pouvez désormais sélectionner des données d'identification IBM Cognos stockées dans IBM SPSS Collaboration and Deployment Services Repository plutôt qu'entrer les informations du nom d'utilisateur et du mot de passe ou d'utiliser une connexion anonyme.

Version 8

IBM SPSS Analytical Decision Management 8 étend la capacité des utilisateurs à déployer l'analyse dans des processus métiers quotidiens, prenant en charge une plage plus large de décisions avec un degré inégalé de connaissances métiers et de contrôle.

- Réglez avec précision et visualisez l'impact combiné des règles, des modèles et de l'optimisation sur des résultats métier avec des fonctions étendus de simulation.
- Résolvez un éventail plus large de problèmes métier avec de nouvelles applications prégénérées et de nouvelles techniques d'optimisation telles que l'optimisation des prix et l'optimisation des assortiments.
- Intégration améliorée avec IBM Operational Decision Management et prise en charge de volumes massifs de données (Big Data) via IBM SPSS Analytic Server.

Une description des principales nouvelles fonctions ajoutées dans la version 8 est fournie ci-dessous :

Nouvelle technologie d'installation. L'installation de produit a été simplifiée et améliorée. Elle utilise maintenant la technologie Installation Manager. Pour plus d'informations, voir le *Guide d'installation d'IBM SPSS Analytical Decision Management*.

Améliorations de l'interface utilisateur. De nombreux changements ont été apportés à l'interface utilisateur pour améliorer la convivialité.

IBM Analytical Decision Management for Predictive Maintenance. IBM SPSS Analytical Decision Management for Predictive Maintenance est une nouvelle application prégénérée qui utilise l'optimisation pour anticiper et éviter des pannes d'équipement avant qu'elles se produisent. Vous pouvez réduire les temps d'arrêt coûteux et utiliser au mieux l'infrastructure et les ressources rares en prévoyant les pannes et en planifiant les actions correctives adéquates au bon moment. Vous anticiperez sans doute d'abord quelles sont les pièces susceptibles de tomber en panne pour les remplacer au moment opportun. La prévision des pannes des machines permet d'optimiser les bénéfices et d'éviter les pertes liées à ces pannes et à l'immobilisation.

IBM Analytical Decision Management for Demand Optimization. IBM SPSS Analytical Decision Management for Demand Optimization est une nouvelle application prégénérée qui utilise l'optimisation pour éviter les excès et les manques de stock coûteux en attribuant la bonne quantité de chaque produit à chaque magasin. Vous pouvez maximiser les profits et éviter les pertes dues aux ruptures de stock et aux frais de réapprovisionnement en affectant la bonne quantité de stock à chaque point de vente.

Améliorations des simulations. Les nouvelles fonctions de simulation combinent la fonctionnalité du bouton **Simuler** précédent (autrefois sur l'onglet Définir) avec la fonctionnalité précédente **Et si ?** (autrefois sur l'onglet Optimiser, Définir les priorités ou Combine) dans une fonction unique. La nouvelle fonctionnalité améliorée de simulation comporte sa propre barre d'outils dans l'angle supérieur droit de l'interface utilisateur, et tient compte des scénarios à enregistrer avec le projet et à conserver d'une session à l'autre.

Déploiement et scoring. Plusieurs améliorations ont été apportées à l'onglet Déployer/Evaluation.

URL de l'application. IBM SPSS Collaboration and Deployment Services a ajouté la possibilité de personnaliser l'URL des applications Web, y compris l'URL de lancement d'IBM SPSS Analytical Decision Management. Contactez votre administrateur pour plus de détails. Pour plus d'informations, voir *IBM SPSS Collaboration and Deployment Services Repository Installation and Configuration Guide* (chapitre sur les racines de contexte d'application).

Améliorations du fichier de statistiques. Les fichiers de données compressés et protégés par mot de passe d'IBM SPSS Statistics (.sav et .zsav) sont désormais pris en charge dans IBM SPSS Analytical Decision Management.

Améliorations du visualiseur de fichiers. Lors de l'affichage des résultats des modèles pris séparément dans IBM SPSS Modeler Advantage, le nom de chaque modèle est un lien qui ouvre une nouvelle boîte de dialogue contenant des informations sur le modèle comme le champ cible et les champs d'entrée. Si le modèle sélectionné est un *modèle d'arbre*, un visualiseur de fichiers s'affiche également. Le visualiseur de fichiers a été amélioré. Vous pouvez désormais effectuer des actions telles qu'un zoom avant et arrière dans l'arborescence, la navigation dans les arborescences volumineuses et le développant ou la réduction de plusieurs noeuds dans l'arborescence.

Options avancées sur l'onglet Optimiser. Pour les applications utilisant une optimisation mathématique complexe (CPLEX) comme IBM SPSS Analytical Decision Management for Campaign Optimization, IBM SPSS Analytical Decision Management for Predictive Maintenance et IBM SPSS Analytical Decision Management for Demand Optimization, des options avancées associées à l'algorithme d'optimisation CPLEX sont maintenant disponibles. Vous pouvez désormais exporter un fichier modèle OPL (langage de programmation d'optimisation) et ajouter les paramètres personnalisés.

Source de vue de données. Un nouveau noeud de *vue de données* a été ajouté à IBM SPSS Modeler 16 pour inclure dans les flux des données qui ont été définies dans une **vue de données analytique** IBM SPSS Collaboration and Deployment Services. La nouvelle source de données de la vue de données peut également être utilisée sur l'onglet Données dans IBM SPSS Analytical Decision Management. Vous pouvez également créer une connexion avec les règles d'IBM Operational Decision Management. Les administrateurs peuvent restreindre l'utilisation de règles complexes qui emploient un modèle de données hiérarchique car les règles de ce type peuvent dégrader les performances.

Avant la version 8, les règles externes nécessitaient des paramètres d'entrée non hiérarchiques correspondant au modèle de données du projet IBM SPSS Analytical Decision Management. Cependant, avec ce nouveau support analytique de vue de données, les règles externes peuvent maintenant comporter des paramètres d'entrée hiérarchiques. Consultez les rubriques «Source Vue de données», à la page 42 et «Verrouillage des options de données», à la page 85 pour plus d'informations et le *chapitre 5* du document *Application Designer's Guide*.

Intégration d'IBM SPSS Analytic Server. IBM SPSS Modeler 16 a introduit IBM SPSS Analytic Server, qui permet d'utiliser HDFS (Hadoop Distributed File System) comme source de données. Les utilisateurs d'IBM SPSS Modeler peuvent créer des flux à l'aide du noeud source HDFS Analytic Server et générer des modèles d'après le système de volumes massifs de données (big data) Hadoop. Les utilisateurs d'IBM SPSS Analytical Decision Management peuvent ouvrir ces flux de modèles dans IBM SPSS Modeler Advantage pour les visualiser (ils ne peuvent pas créer de projets à l'aide de sources de données Analytic Server).

Nouveau chapitre du document Application Designer's Guide. Un chapitre a été ajouté au document *IBM SPSS Analytical Decision Management Application Designer's Guide*. Il fournit un exemple de configuration d'une application d'optimisation personnalisée. L'application IBM SPSS Analytical Decision Management for Campaign Optimization utilise un fichier modèle OPL (langage de programmation d'optimisation) nommé *CampaignOptimization.mod*. Dans certains cas, l'application standard d'optimisation de campagne et son modèle OPL ne sont pas adaptés aux buts du problème métier actuel. Le nouveau chapitre parcourt un exemple de modèle OPL personnalisé généré à l'aide d'IBM ILOG CPLEX Optimization Studio et une application d'optimisation IBM SPSS Analytical Decision Management personnalisée pour une société fictive.

Résultats de l'évaluation sous forme de tableau. Une nouvelle option d'**inclusion des résultats sous forme de tableau** est disponible lors des évaluations. Si vous désélectionnez l'option, les performances s'en trouvent améliorées car cela prend plus de temps d'exécuter des évaluations incluant les résultats sous forme de tableau.

Nouveau paramètre pour le nombre maximal de champs nominaux. Un nouveau paramètre a été ajouté à la boîte de dialogue Propriétés pour définir le nombre maximal de membres autorisés dans les champs de données *Nominal*. Cette fonction est utilisée sur l'onglet Données. Pour plus d'informations, voir «Propriétés», à la page 33.

Connexion unique. Les connexions ODBC peuvent désormais utiliser la connexion unique pour les connexions de base de données IBM SPSS Modeler Server et ODBC avec IBM SPSS Analytical Decision Management. Certaines étapes de configuration sont nécessaires. Consultez la rubrique de remarques sur la sécurité de la documentation d'installation.

Processus SQL push-back. Lorsque les données source proviennent d'une base de données, l'exécution est renvoyée par commande push à la base de données si possible (dans un processus connu sous le nom de "SQL push-back"). Généralement, IBM SPSS Analytical Decision Management prend en charge le même niveau de SQL push-back qu'IBM SPSS Modeler, mais avec les limitations suivantes :

- Si le projet inclut des champs ou des règles de sélection supplémentaires avec des références à des modèles ou à des règles externes, ces références ne sont pas renvoyées à la base de données par commande push. Cela peut diminuer le nombre d'actions réalisables par ce biais pour le flux de projet dans son ensemble.
- Dans les applications autres qu'IBM SPSS Modeler Advantage, la fonction de renvoi par commande push est activée uniquement pour les règles de sélection (sélections globales et sélections de membre pour la première dimension). Le degré de renvoi par commande push peut être accru si vous sélectionnez un champ Entier comme ID d'entité dans le modèle de données du projet.

Version 7.0.0.1

IBM SPSS Decision Management 7 groupe de correctifs 1 (7.0.0.1) a ajouté les fonctions suivantes :

Présentation des données. Une nouvelle fonction de présentation des données est disponible sur l'onglet Données de toutes les applications, sur les onglets Modélisation et Evaluation d'IBM SPSS Modeler Advantage, ainsi que sur l'onglet Déployer des applications qui incluent la fonction *Evaluer maintenant*. Pour plus d'informations, voir la rubrique «Présentation des données», à la page 45. Cette fonction est disponible uniquement avec IBM SPSS Modeler Server 15 groupe de correctifs 1 (15.0.0.1). Cette fonction est indisponible dans la version 15.

Rapport récapitulatif d'évaluation. Après l'évaluation d'un projet, une nouvelle section *Rapport récapitulatif d'évaluation* fournit des détails tels que l'emplacement vers lequel les données ont été évaluées, la taille du fichier (le cas échéant), le nom de la base de données et de la table (le cas échéant), le nom du modèle ou du projet, nombre d'enregistrements évalués, et, dans certains cas, un graphique et/ou un tableau récapitulatifs du modèle. La section Rapport récapitulatif d'évaluation est disponible dans IBM SPSS Modeler Advantage et sur l'onglet Déployer des applications qui incluent la fonction *Evaluer maintenant*. La taille de fichier ne s'affiche que sur IBM SPSS Modeler Server 15 groupe de correctifs 1 (15.0.0.1). Cette fonction est indisponible dans la version 15.

Télécharger une copie locale du fichier scoré. Si vous choisissez d'envoyer les données de sortie scorées vers un fichier à plat, vous pouvez également télécharger une copie locale de ce fichier. Les administrateurs peuvent utiliser l'outil basé sur navigateur, IBM SPSS Collaboration and Deployment Services Deployment Manager, pour activer ou désactiver cette fonction, ainsi que pour augmenter ou réduire la taille de fichier maximale autorisée. Si le bouton **Télécharger le fichier** est inactif, cela signifie que la taille de fichier dépasse la limite maximale définie par l'administrateur ou que celui-ci a désactivé la fonction. La fonction est désactivée par défaut. Cette fonction est disponible uniquement avec IBM SPSS Modeler Server 15 groupe de correctifs 1 (15.0.0.1). Cette fonction est indisponible dans la version 15.

Nouveaux paramètres facultatifs pour la génération de modèles. Lors de la génération d'un modèle dans IBM SPSS Modeler Advantage, vous disposez de nouveaux paramètres facultatifs, dont des options de partitionnement de données, le choix des techniques (algorithmes) de modélisation à inclure lors de la génération du modèle, la spécification du nombre maximal de modèles à conserver et à combiner, la définition des critères de profit avant de générer le modèle. Les options disponibles dépendent du type de modèle et de la cible sélectionnée. Pour plus d'informations, voir la rubrique «Paramètres facultatifs de modèle», à la page 65. Ces nouvelles fonctions sont disponibles uniquement avec IBM SPSS Modeler Server 15 groupe de correctifs 1 (15.0.0.1). Ces fonctions sont indisponibles dans la version 15.

- Sélection des techniques de modélisation à inclure lors de la génération du modèle
- Définition des critères de profit
- Visualisation des détails de haut niveau de modèles individuels
- Visualisation des détails supplémentaires d'un modèle sélectionné

Visualiseur de fichiers. Lorsque vous avez généré un modèle dans IBM SPSS Modeler Advantage et visualisé ses résultats, vous pouvez cliquer sur son nom. Le nom de chaque modèle est un lien qui ouvre une nouvelle boîte de dialogue contenant des informations sur le modèle (telles que le champ cible et les champs d'entrée). Si le modèle sélectionné est un *modèle d'arbre*, désormais un visualiseur de fichiers s'affiche également. Le visualiseur de fichiers est disponible uniquement avec IBM SPSS Modeler Server 15 groupe de correctifs 1 (15.0.0.1). Cette fonction est indisponible dans la version 15.

Améliorations de l'évaluation (scoring). Diverses améliorations ont été apportées à la boîte de dialogue Evaluer dans IBM SPSS Modeler Advantage. Par exemple, un nouveau bouton **Modifier la sélection de modèle** est disponible pour modifier les modèles à inclure dans l'évaluation. Vous pouvez également copier les sélections de modèle en cours effectuées pour les replacer dans la table de résultats du modèle individuel. Certaines nouvelles options sont disponibles uniquement pour des types de modèle spécifiques (la plupart des nouvelles options s'appliquent uniquement au type de modèle *Prédictif* ; certaines sont également destinées au type de *modèle de règle prédictif*). Les nouvelles fonctions suivantes de la boîte de dialogue Evaluer sont activées uniquement avec IBM SPSS Modeler Server 15 groupe de correctifs 1 (15.0.0.1). Ces fonctions sont indisponibles dans la version 15.

- Graphiques de performances pour une cible numérique
- Simulation de profit pour une cible numérique
- Graphiques de matrice de coïncidence (pour toutes les cibles)
- Vue Table de tous les graphiques
- Bouton **Modifier la sélection de modèle**
- Option **Inclure les évaluations de modèle individuelles pour les graphiques de performances**

Microsoft Windows 8 et Internet Explorer 10. IBM SPSS Collaboration and Deployment Services 5 groupe de correctifs 1 (5.0.0.1) ajoute le support de Windows 8 et d'Internet Explorer 10. Ces logiciels ne sont pas pris en charge dans IBM SPSS Collaboration and Deployment Services 4.2.1 ou 5.

Améliorations de l'interface utilisateur. De nombreux légers changements ont été apportés à l'interface utilisateur pour améliorer la convivialité.

Version 7

IBM SPSS Decision Management 7 (7.0.0.0) a ajouté les fonctions suivantes :

Optimisation. L'optimisation vise à identifier la solution qui répond le mieux à un objectif précis, comme la maximisation des revenus d'une campagne marketing ou la réduction des risques de fraude ou de résiliation. Le résultat de l'optimisation est une solution qui répond à cette question à travers toutes les combinaisons possibles au sein de votre population, de la manière qui maximise le mieux la valeur que vous souhaitez optimiser (par exemple, le profit).

Modèles de cluster (classification). Nouveauté d'IBM SPSS Modeler Advantage, les méthodes de modélisation de classe permettent de découvrir des groupes d'enregistrements de données présentant des valeurs ou des motifs similaires. Ces techniques sont utilisées dans le marketing (segmentation de la clientèle) et dans d'autres applications métier (les enregistrements placés dans des classes à enregistrement unique peuvent contenir des erreurs ou des cas de fraude). Les modèles de cluster sont souvent appelés modèles d'apprentissage non supervisé car ils n'utilisent pas de champ cible et ne renvoient pas de prévision spécifique pouvant être évaluée comme vraie ou fausse. IBM SPSS Decision Management permet de générer automatiquement des modèles de cluster. Vous pouvez également combiner, renommer, modifier, supprimer des clusters ou ajouter des clusters personnalisés d'après la connaissance métier antérieure.

Modèles d'association. Nouveauté d'IBM SPSS Modeler Advantage, le modèle d'association permet de prédire quels éléments ont le plus de chance d'apparaître ensemble ainsi que l'intensité de leur relation. Les règles d'association associent une conclusion particulière (l'achat d'un produit particulier, par exemple) à un ensemble de conditions (l'achat de plusieurs autres produits, par exemple).

Données. L'onglet Données permet désormais de disposer de plusieurs entrées de données, de définir le modèle de données du projet avec une vue d'application, de se référer aux sorties du modèle comme s'il s'agissait de données, de télécharger des fichiers de données locales, ainsi que d'obtenir l'aperçu et le comptage des données.

Améliorations de l'intégration de données d'IBM Cognos. Comme avec IBM SPSS Modeler 15, IBM SPSS Decision Management prend désormais en charge l'importation de rapports Cognos pré-générés, les filtres et les calculs Cognos ainsi que le contrôle de la désagrégation automatique.

Éléments multiples. Les applications peuvent désormais inclure plusieurs hiérarchies de dimension. Lors du lancement d'une application, si l'application sélectionnée est configurée pour inclure plusieurs éléments, vous serez invité à sélectionner un ou plusieurs des éléments disponibles (hiérarchies) pour les utiliser dans l'application. Sur l'onglet Définir, l'arborescence des dimensions indique la gamme de décisions ou de recommandations pouvant être renvoyées. Si une application comprend plusieurs hiérarchies de dimension, celles-ci sont incluses sur des onglets séparés dans la section de l'arborescence des dimensions située sur l'onglet Définir.

Allocation dynamique. Dans les éditions précédentes, un élément ne pouvait être alloué que s'il existait dans l'arborescence des dimensions. Désormais, vous pouvez lire les éléments de dimension sans enfant dans une table liée par une clé sur le modèle de données du projet. Par exemple, soit un élément de dimension sans enfant appelé *Véhicules*. Vous pouvez définir des allocations dynamiques sur *Véhicules* pour référencer des champs dans une table (par exemple, *BMW*, *Audi* et *Ford*).

Allocations multiples. Une nouvelle option, **Allocations multiples**, permet d'allouer des enregistrements à plusieurs éléments de dimension (par exemple, à plusieurs offres) pour chaque segment. Vous pouvez définir plusieurs allocations par règle, plutôt qu'une seule.

Améliorations de l'interface utilisateur. De nombreux changements et ajouts ont été apportés à l'interface utilisateur d'IBM SPSS Decision Management afin d'améliorer la convivialité et l'évolutivité de toutes les applications.

Améliorations de la galerie. Diverses améliorations ont été apportées à la galerie (par exemple, l'option **Vider la galerie**) qui permet de supprimer plusieurs éléments de galerie à la fois).

Améliorations de l'éditeur de règles. Selon l'emplacement à partir duquel l'éditeur de règles est lancé, les utilisateurs peuvent désormais définir le type des règles de sélection (Inclure ou Exclure), affecter des points de risque pour les règles d'agrégation, sélectionner une dimension pour les règles d'allocation et utiliser les contrôles *OU* et *Fractionner OU* pour combiner plusieurs expressions d'une règle et créer une expression *OU*. Lors de la création de règles, vous pouvez désormais sélectionner une dimension dans une arborescence hiérarchique (pour les applications comportant plusieurs arborescences) ou lancer l'éditeur d'expression pour créer une expression. Auparavant, cette fonctionnalité n'était pas disponible dans la boîte de dialogue de l'éditeur de règle.

Intégration étendue d'IBM Operational Decision Management. Dans IBM SPSS Collaboration and Deployment Services Deployment Manager, vous pouvez utiliser la nouvelle boîte de dialogue Ajouter une règle externe pour définir des références à des jeux de règles IBM Operational Decision Management utilisables ensuite dans IBM SPSS Decision Management. Auparavant, ces règles n'étaient accessibles que par l'intermédiaire d'un service Web. Dans cette édition, vous pouvez désormais faire référence à des règles locales dans la base de données IBM Operational Decision Management. Pour la nouvelle option locale, la règle est exécutée via IBM SPSS Modeler Server (le même serveur défini dans Deployment Manager utilisé par IBM SPSS Decision Management).

Améliorations de WhatIf?. Diverses améliorations ont été apportées à la fonctionnalité disponible dans la boîte de dialogue WhatIf?. Par exemple, pour la date de simulation, vous pouvez désormais entrer la plage de dates à utiliser pour effectuer l'évaluation WhatIf?.

Télécharger les fichiers de flux. Vous pouvez désormais ouvrir tout fichier de flux IBM SPSS Modeler à partir du système de fichiers local et le stocker dans le référentiel, si vous le souhaitez.

Données d'identification IBM SPSS Modeler Server multiples. Si plusieurs ensembles de données d'identification IBM SPSS Modeler Server sont définis dans IBM SPSS Collaboration and Deployment Services Deployment Manager, vous pouvez désormais en indiquer également plusieurs à utiliser avec IBM SPSS Decision Management. Ensuite, lorsqu'un utilisateur IBM SPSS Decision Management se connecte, le serveur essaie le premier ensemble de la liste afin de voir si l'utilisateur correspondant possède les droits requis, puis il procède de même avec les autres, en suivant l'ordre de la liste, jusqu'à ce qu'il en trouve un qui possède ces droits. Si l'utilisateur n'a de droits sur aucun des ensembles de données d'identification de la liste, il ne peut pas utiliser IBM SPSS Decision Management. Cela permet de mieux contrôler l'accès des utilisateurs aux différentes parties du système de fichiers IBM SPSS Modeler Server.

Version 6.2

IBM SPSS Decision Management 6.2 a ajouté les fonctions suivantes :

Support des données Cognos. Utilisé avec IBM SPSS Modeler Server 14.2 ou version ultérieure, IBM SPSS Decision Management 6.2 a ajouté la possibilité de lire des données d'un serveur IBM Cognos Business Intelligence lors de la définition de sources de données et de renvoyer des résultats d'évaluation à un serveur Cognos BI. Une fois que vous avez défini une connexion au serveur Cognos BI, vous pouvez parcourir les packages, les dimensions, les requêtes et les rapports pour sélectionner les champs souhaités. Pour plus d'informations, voir la rubrique «Source IBM Cognos BI», à la page 42.

Intégration du système de gestion de règles métier ILOG Des règles créées dans un système de gestion de règles métier tel qu'ILOG peuvent être référencées et utilisées dans des applications IBM SPSS Decision Management, si elles ont été développées de manière à prendre en charge le même modèle de données que celui utilisé dans le projet IBM SPSS Decision Management en cours. Ces règles externes peuvent être sélectionnées et utilisées de la même manière que d'autres modèles et règles dans

l'application IBM SPSS Decision Management. Pour plus d'informations, voir la rubrique «Utilisation de règles IBM Operational Decision Management», à la page 57. ILOG est désormais désigné "IBM Operational Decision Management".

Support de Linux sur System Z. SUSE Enterprise Server 10 est pris en charge sur IBM System z10 64 bits.

Chapitre 2. Exemple de modélisation des réponses des clients

IBM SPSS Modeler Advantage

IBM SPSS Modeler Advantage est une application facile d'emploi qui met à la disposition des utilisateurs métier toute la puissance de la modélisation prédictive. A l'aide de modèles prédictifs, vous pouvez identifier des motifs basés sur ce qui s'est produit dans le passé, et les utiliser pour prévoir ce qui est susceptible de se produire à l'avenir.

Par exemple, vous pouvez utiliser un modèle pour prévoir quels clients sont les moins susceptibles de résilier leur contrat ou les plus susceptibles de répondre à une offre particulière, d'après des caractéristiques telles que les revenus, l'âge ainsi que les organisations ou groupes auxquels ils appartiennent. Les prévisions obtenues peuvent être utilisées pour générer des listes de clients cible ou des cas présentant un intérêt, tels que l'entrée d'une planification stratégique, où être intégrées avec des règles dans le contexte d'une application prédictive.

Optimisation du profit pour les campagnes de publipostage

Dans cet exemple de modèle prédictif, vous commercialisez des produits et des services financiers pour les clients d'une banque fictive. Dans une campagne de test récente, un certain nombre de clients se sont vu proposer l'un des nouveaux services. Leurs réponses ont été enregistrées. Toutefois, les taux de réponse ont été inférieurs aux résultats attendus et la campagne n'a pas pu générer de profit.

Grâce à IBM SPSS Modeler Advantage, votre but est de convertir cette perte en profit en ciblant les clients les plus appropriés en fonction de vos coûts et revenus attendus, et d'autres facteurs.

Les résultats de la campagne de test sont résumés ci-dessous.

Tableau 1. Résultats de la campagne de test

Résultats de la campagne de test	
Nombre de courriers	3000
Nombre de réponses	111
Taux de réponse	3,7 %
Coût	6750
Revenus	6660
Profit total (perte)	(90)

Cet exemple utilise les fichiers de données *bank_customer_data.txt* et *bank_response_data.txt*, distribués et installés avec IBM Analytical Decision Management. Une version complète de cet exemple est également fournie dans le fichier *bank_response_model.str*. Contactez votre administrateur pour en savoir plus sur l'installation et pour obtenir des exemples de fichiers si nécessaire. Pour plus d'informations, voir la rubrique «Exemples de fichiers».

Exemples de fichiers

Des données et autres exemples de fichiers sont fournis avec IBM Analytical Decision Management. Ils peuvent être utilisés pour reproduire les exemples donnés dans les guides d'utilisation.

Données

Lors de l'installation, différents exemples de fichiers de données s'installent dans le répertoire IBM SPSS Modeler Server\Demos (par exemple, C:\Program Files\IBM\SPSS\ModelerServer\18\Demos). Ces données peuvent être utilisées avec les exemples d'applications prégénérées et comprennent les fichiers suivants :

- *bank_customer_data.txt*
- *bank_response_data.txt*
- *customer_contact_history.txt*
- *customer_summary_data.txt*
- *insurance_claim_data.txt*
- *insurance_fraud_data.txt*
- *inventory_orders.txt*
- *inventory_stock.txt*
- *machine_log_data.txt*
- *machine_summary_data.txt*
- *retail_purchase_data.txt*

Les exemples de règles, de modèles et de fichiers d'application qui utilisent ces fichiers de données adoptent l'emplacement par défaut indiqué ci-dessus. Si les fichiers de données sont installés dans une autre unité ou à un autre emplacement, les chemins de source de données doivent être mis à jour dans les exemples d'applications en conséquence.

Exemples

Différents exemples de règles, de modèles et de fichiers d'application sont fournis pour être utilisés avec les applications IBM Analytical Decision Management prégénérées. Ces exemples sont fournis dans un fichier téléchargeable. Pour importer des exemples de fichiers d'une ou plusieurs applications prégénérées dans IBM SPSS Collaboration and Deployment Services Repository, procédez comme suit.

1. Téléchargez le fichier *ADM_18_Demo.zip* dans un emplacement temporaire et extrayez-le.
2. Ouvrez IBM SPSS Collaboration and Deployment Services Deployment Manager et ouvrez votre connexion au serveur de contenu.
3. Cliquez avec le bouton droit de la souris sur le dossier du référentiel de contenu et sélectionnez **Importer**.
4. Dans la boîte de dialogue Importer, cliquez sur **Parcourir** et sélectionnez le fichier *DecisionManagementExamples pes* que vous avez précédemment copié ou extrait dans un emplacement temporaire à l'étape 1. Cliquez sur **OK**.
5. Dans la boîte de dialogue Import Conflicts, cliquez sur **OK**. Une fois l'importation terminée, vous verrez apparaître les nouveaux dossiers ci-dessous ; ils contiennent des exemples de projets que vous pourrez ouvrir dans IBM Analytical Decision Management.

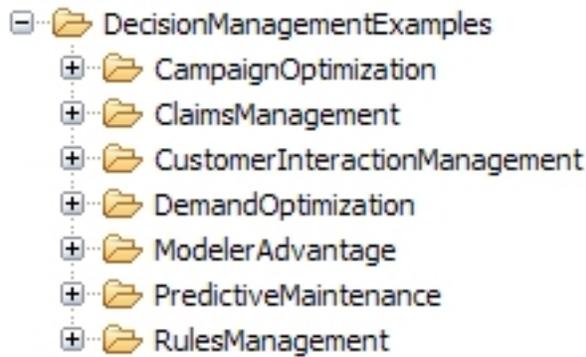


Figure 1. Dossiers destinés aux exemples d'artefacts

Vous remarquerez tous les exemples de fichiers, de modèles et de règles d'application partent du principe que les fichiers de données sont installés dans le répertoire *Demos* d'IBM SPSS Modeler Server (par exemple, *C:\Program Files\IBM\SPSS\ModelerServer\18\Demos*). Si les fichiers de données sont installés à un autre emplacement, les chemins de source de données doivent être mis à jour dans les exemples d'applications en fonction de la configuration locale.

Lancement de l'application

La page de lancement d'IBM Analytical Decision Management comporte des liens vers les applications disponibles sur votre site. Ces liens peuvent être configurés par l'administrateur ; vous pouvez certainement sélectionner des applications supplémentaires en cliquant sur le lien **Ajouter une application** qui apparaît dans l'angle supérieur droit.

1. Pour utiliser l'exemple, ouvrez la page de lancement d'IBM Analytical Decision Management. Pour faciliter les accès ultérieurs, vous pouvez ajouter dans votre navigateur un signet vers cette page ou créer un raccourci sur votre bureau électronique.
2. Pour créer un modèle, sélectionnez **Nouveau** dans la liste déroulante de l'application du groupe IBM SPSS Modeler Advantage, puis cliquez sur **Atteindre**. La page d'accueil d'IBM SPSS Modeler Advantage s'affichera.



Figure 2. Page d'accueil d'IBM SPSS Modeler Advantage

Chaque icône de la page d'accueil représente une étape différente dans le processus de modélisation, depuis la définition des données jusqu'à la création et l'évaluation du modèle. Pour commencer, cliquez sur l'icône Données à gauche.

Configuration des sources de données

L'onglet Données permet de configurer des sources de données pour les exemples de fichier de données utilisé dans l'exemple. Les ensembles de données peuvent être configurés par l'administrateur. Vous pouvez également ajouter les vôtres.

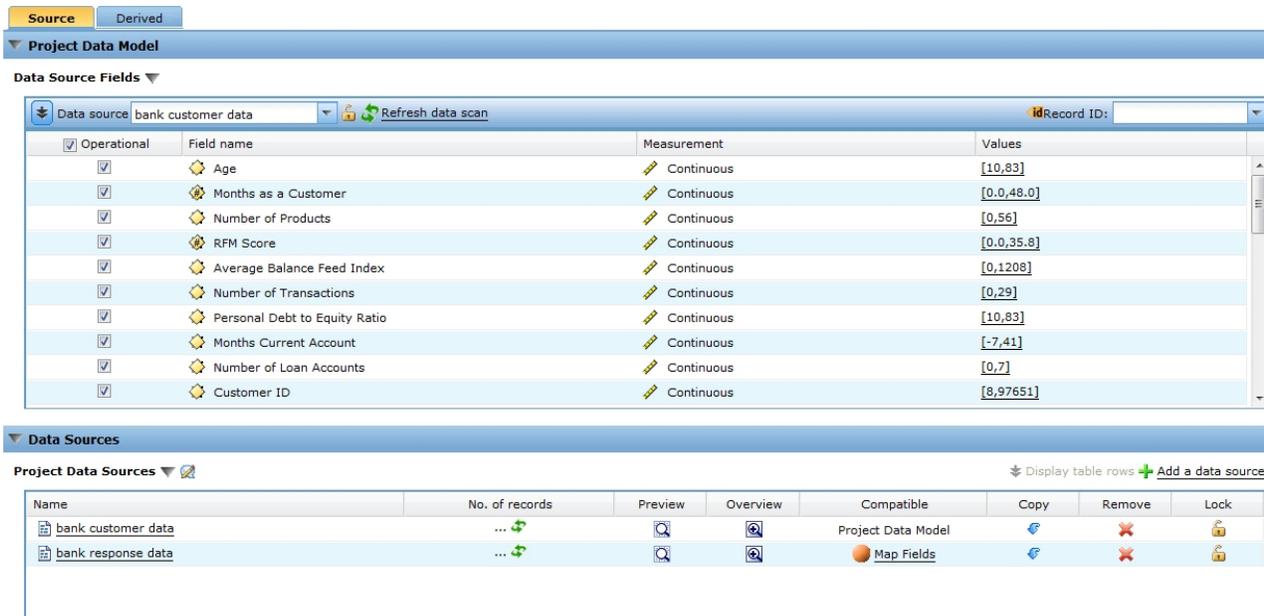


Figure 3. Data tab

Le modèle de données du projet définit les champs disponibles dans toute l'application et peut être nécessaire pour les règles et les modèles à évaluer (scorer). Toutes les autres sources de données doivent être mappées au modèle de données pour être utilisées par l'application.

L'exemple ci-après utilise deux fichiers de données, distribués avec l'application :

- **bank_customer_data.txt** inclut des informations sur les clients actuels (dont l'identificateur de client, l'historique des paiements, le nombre de produits détenus, le nombre de mois de fidélité), ainsi que des informations démographiques (comme l'âge et le revenu). Le modèle de données du projet est basé sur ce fichier.
- **bank_response_data.txt** contient des informations sur les clients qui ont bénéficié d'un ou de plusieurs services lors de campagnes précédentes et enregistre leurs réponses. Par l'analyse de ces données, l'application peut identifier les clients qui sont davantage susceptibles de répondre à des offres similaires lors de campagnes ultérieures.

Par défaut, les exemples de fichiers de données sont installés dans le dossier IBM SPSS Modeler Server Demos, accessible à partir de votre application (par exemple, C:\Program Files\IBM\SPSS\ModelerServer\18\Demos\). Si nécessaire, contactez l'administrateur pour obtenir une aide à l'installation de ces fichiers.

Pour définir le modèle de données du projet :

1. Sur le panneau Sources de données du projet, sélectionnez **Ajouter une source de données** dans le menu déroulant **Source de données**.
2. Indiquez *bank_customer_data* (données bancaires des clients) comme nom de source de données.
3. Sélectionnez **Fichier** et *bank_customer_data.txt*.
4. Sous **Indiquer les champs d'entrée**, vérifiez que le type de mesure pour le champ *Has Bad Payment Record* (a un enregistrement de paiement incorrect) est défini sur **Indicateur**. Le paramètre Indicateur spécifie que le champ a deux valeurs possibles, *true* (vrai) ou *false* (faux), généralement codées respectivement *T* et *F* (ou 0 et 1 dans un champ numérique), pour signaler la présence ou l'absence de l'attribut.
5. Sous **Indiquer les champs d'entrée**, vérifiez également que le type de mesure pour le champ *Gender* (genre) est défini sur **Nominal**. (Les valeurs de ce champ indiquent simplement le genre, masculin ou féminin, contrairement à l'indicateur qui fournit *true* ou *false*).

6. Cliquez sur **Liaisons**. La boîte de dialogue Liaisons s'affiche.
7. Cliquez sur **Enregistrer**.

Etant donné qu'il s'agit de la première source de données définie pour le projet, elle est automatiquement sélectionnée pour le modèle de données du projet et utilisée pour définir les champs requis par l'application. Tous les autres ensembles de données sont mappés par rapport à cette source.

Pour définir la source de données de réponse :

L'ensemble de données de réponse contient de nombreux champs identiques à ceux utilisés dans le modèle de données du projet, ainsi que deux champs supplémentaires nommés *Campaign ID* (ID campagne) et *Response* (Réponse). Ces champs, qui indiquent l'offre proposée aux clients et leur réponse éventuelle, peuvent être utilisés pour générer un modèle prédictif sur les clients davantage susceptibles de répondre à l'avenir.

1. Sur le panneau Sources de données du projet, sélectionnez **Ajouter une source de données** dans le menu déroulant **Source de données**.
2. Indiquez *bank response data* (données bancaires de réponse) comme nom de source de données.
3. Sélectionnez **Fichier** et sélectionnez *bank_response_data.txt*.
4. Sous **Indiquer les champs d'entrée**, vérifiez également que le type de mesure pour le champ *Response* (réponse) est défini sur **Indicateur**.
5. Sous **Indiquer les champs d'entrée**, décochez la case en regard du champ *Campaign ID* (ID campagne) pour indiquer que le champ concerné ne doit pas être utilisé en entrée lors de la modélisation. (Ce champ peut être employé pour indiquer les campagnes pour lesquelles un client a été sélectionné dans le passé, mais il n'est pas utile pour prédire les réponses ultérieures.)
6. Cliquez sur **Enregistrer**.

Etant donné que certains champs du modèle de données sont introuvables dans l'ensemble de données, une icône orange s'affiche pour la source de données concernée dans la colonne **Compatible** de l'onglet Données, indiquant que quelques champs correspondent, mais pas tous. L'ensemble de données est approprié pour la modélisation, mais ne convient pas pour la simulation, l'évaluation (scoring) ou le déploiement. (Pour être utilisables pour ces fonctions, tous les champs du modèle de données doivent être présents.)

Génération du modèle

L'onglet Modélisation vous permet d'indiquer la source de données, le champ cible ainsi que d'autres options pour créer et évaluer le modèle.

Model type
 Predictive model Change Model...

Data source
 bank response data + +

Records used to build: ... ↻

Target 🎯
 Response

Evaluate
Test

Optional Settings ▶ ▶ Build Model

▼ Predictive Model Results

Display: Combined model results Individual model results

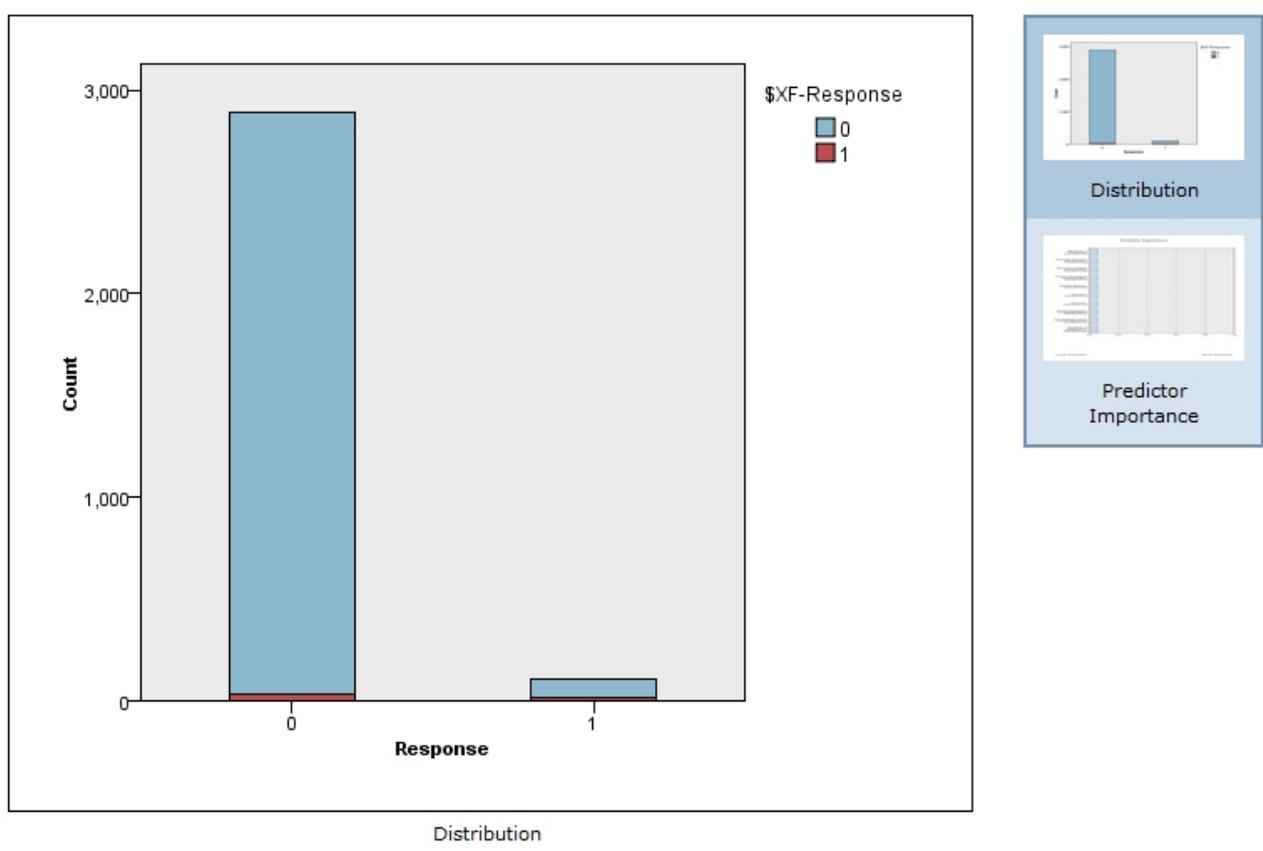


Figure 4. Model tab

1. Sélectionnez les *données bancaires de réponse* pour la source de données, puis *Réponse* comme champ cible.
2. Si le message **Analyse de données requise** s'affiche, cliquez sur l'icône pour exécuter une analyse de données. Cliquez ensuite sur **Générer un modèle**.

Lorsque le modèle est terminé, un graphique de fréquences s'affiche. Le graphique présente les valeurs observées sur l'axe horizontal, la valeur 1 indiquant une réponse positive.

- La hauteur des barres reflète le taux de réponse global. Sur les 3 000 clients que compte l'ensemble de données de réponse, 111 ont répondu, pour un taux d'occurrences global d'environ 3,8 %.
- Les zones colorées dans chaque barre représentent les réponses prévues par le modèle, le rouge indiquant une prévision positive et le bleu indiquant une réponse négative, comme indiqué dans la

légende dans le coin supérieur droit. En raison du faible taux de réponse global, il y a relativement peu de réponses prévues (zones rouges). Cependant, la barre fausse (fausse) (*réponse = 0*) est principalement bleue, indiquant que la majorité des clients pour lesquels aucune réponse n'était attendue n'ont effectivement pas répondu.

Pour avoir une meilleure idée de la façon dont s'exécute le modèle, passez à la section suivante («Évaluation du modèle»).

Evaluation du modèle

Dans l'onglet Modélisation, cliquez sur **Évaluer** pour voir les performances du modèle lorsque celui-ci est appliqué à vos données clients ainsi que pour comprendre l'impact que l'utilisation du modèle peut avoir sur votre résultat net.

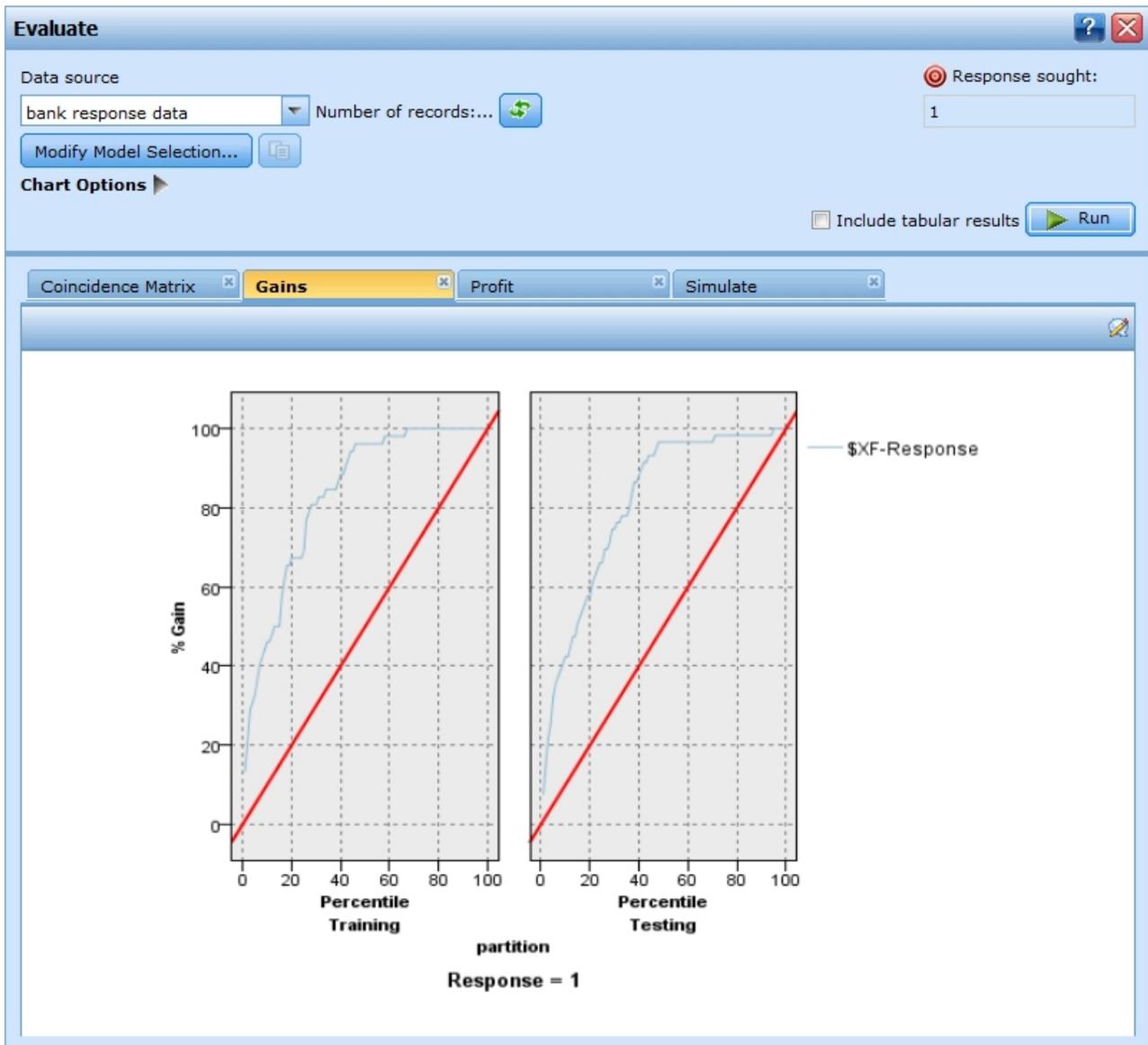


Figure 5. Evaluating the model

1. Pour la source de données, sélectionnez les *données bancaires de réponse*.
2. Sélectionnez les types de graphique à générer suivants : **Graphique de gain** et **Graphique de profits**.

3. Sélectionnez **Simuler le profit maximum** et indiquez *3000* pour la population (soit le nombre de votre campagne de test).
4. Indiquez *2,25* pour le coût et *60* pour les revenus. Ces valeurs indiquent le coût que vous aurez à supporter pour chaque proposition faite (coûts de production, de traitement et de mailing entre autres), ainsi que les revenus attendus pour chaque réponse favorable.
5. Pour **Réponse recherchée**, sélectionnez *1*.
6. Cliquez sur **Exécuter**.

Le graphique de gain présente les améliorations ou "lift" pouvant être obtenu à l'aide du modèle, comme indiqué dans la version précédente. Etant donné que l'option **Partitionner automatiquement les données** a été sélectionnée (sous Paramètres facultatifs) lors de la génération du modèle, des graphiques distincts sont affichés pour les partitions d'apprentissage et de test. Chaque graphique affiche la proportion d'occurrences (réponses positives) à chaque percentile par rapport au nombre total d'occurrences dans l'échantillon. La diagonale plate trace la réponse attendue si aucun modèle n'est utilisé. Pour atteindre la moitié des occurrences, vous devrez contacter la moitié des personnes. La courbe indique l'amélioration à l'aide du modèle. En visant les 60 % de clients les plus susceptibles de répondre, vous pouvez atteindre la quasi-totalité des occurrences avec un mailing réduit.

Le graphique de profits affiche les gains en termes de profit global, en fonction des valeurs de coût et de revenu indiquées. Dans l'échantillon utilisé pour configurer le modèle, le profit atteint son maximum juste en dessous du 30e percentile.

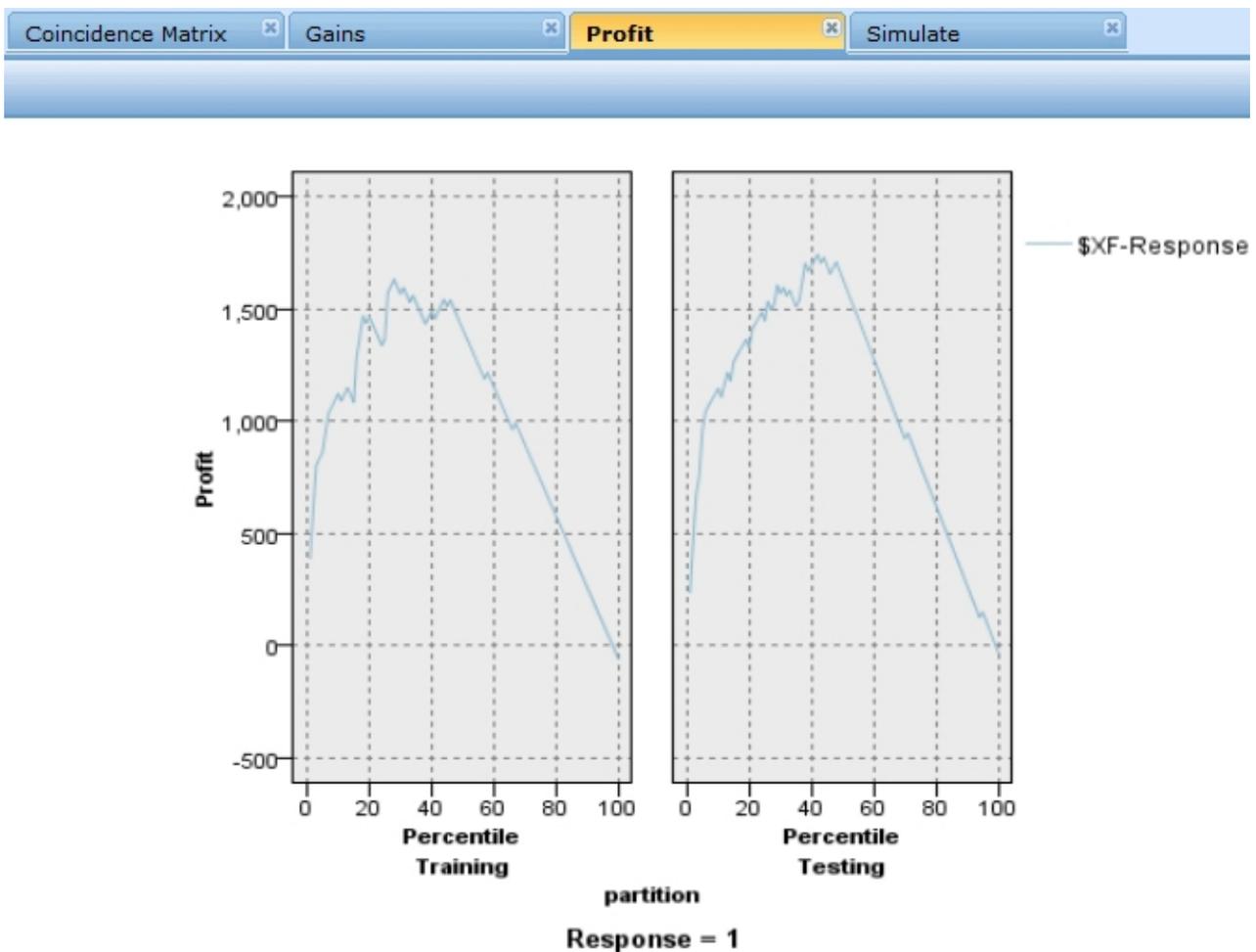


Figure 6. Profit chart

Pour comprendre le fonctionnement, accédez à l'onglet Simuler pour voir le profit maximal attendu correspondant à la taille de la population que vous avez indiquée.

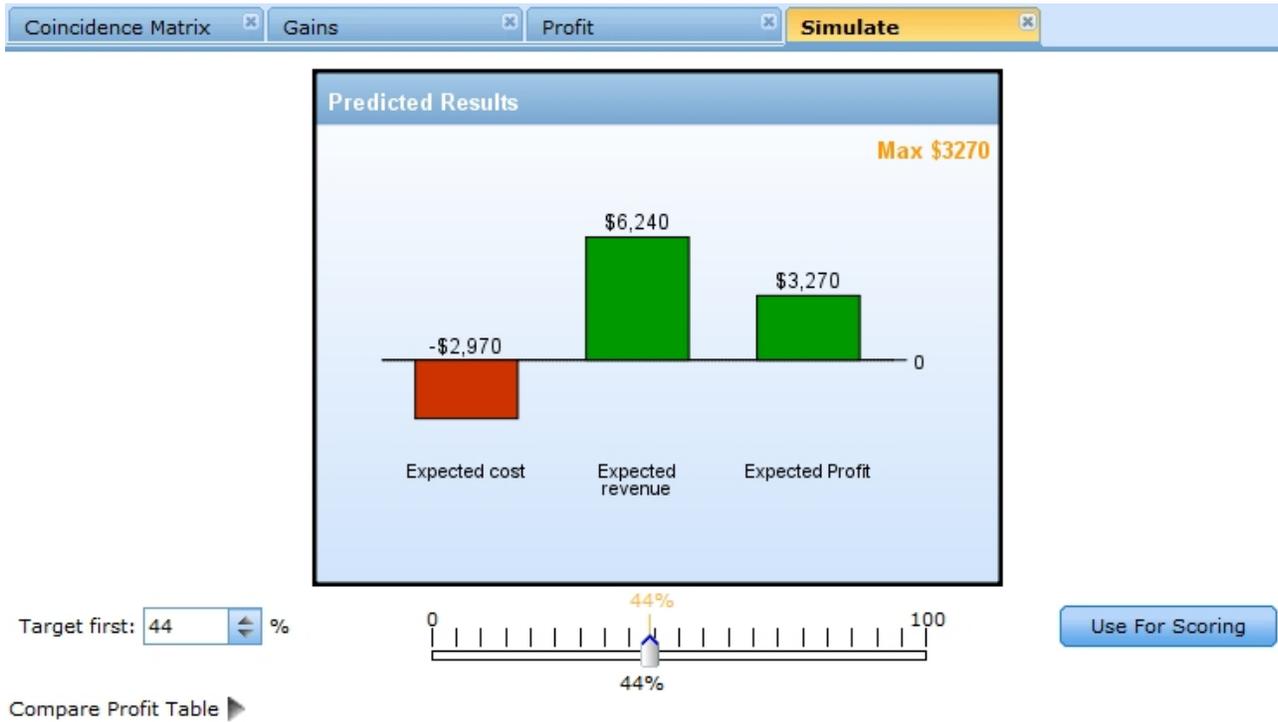


Figure 7. Profit simulation

Le profit attendu est calculé à l'aide de la formule suivante :

$$\text{profit attendu} = (\text{probabilité de réponse} * \text{revenu}) - \text{coût}$$

Où la probabilité de réponse est déterminée pour chaque client en fonction du modèle.

En ciblant les clients identifiés par le modèle comme étant les plus susceptibles de répondre, vous pouvez optimiser votre profit par client. En incluant davantage de clients, le profit total augmentera jusqu'à un certain point, même si le retour par client diminue graduellement en raison de l'ajout de clients avec des taux de réponse inférieurs. La simulation vous aide à décider où poser des limites pour obtenir le maximum de profit.

Compare Profit Table ▼

	Targeted 44%	Random 44%	Everyone
Offers made	1320	1320	3000
Expected response	104	48.8	111
Expected response(%)	7.9	3.7	3.7
Expected cost	\$2970	\$2970	\$6750
Expected revenue	\$6240	\$2930.4	\$6660
Expected Profit	\$3270	\$-39.6	\$-90

Figure 8. Profit comparison

La table Comparer la table de profit détaille ce résultat.

- **44 % ciblés.** Cette colonne affiche le résultat lorsque 44 % des clients présentant les taux les plus élevés d'après les prévisions du modèle sont ciblés. En incluant de façon sélective les 1 320 clients les plus à même de répondre, vous pouvez optimiser votre retour avec un profit attendu de 3 270 euros.
- **44 % aléatoire.** Affiche le résultat pour un échantillon aléatoire de la même taille. Dans ce cas, le nombre d'offres envoyé est le même (1 320), mais la réponse attendue est inférieure (conforme au taux de la campagne de test) car vous ne tirez pas parti des informations supplémentaires fournies par le modèle. Il en découle une perte car le taux de réponse global n'est pas suffisant par rapport au coût de mailing des offres.
- **Tous.** Affiche le résultat lorsque l'échantillon inclut les 3 000 clients. Le résultat est identique au résultat renvoyé par votre campagne de test. Là également, il en découle une perte car vous ne tirez pas parti du modèle.

Au vu des résultats, vous pouvez constater que l'utilisation du modèle permet d'améliorer le taux de réponse afin de transformer une campagne improductive en campagne rentable. La différence entre une perte de 90,00 euros et un profit de 3 277,50 euros illustre l'avantage que permet la modélisation prédictive. Ce retour se fonde sur les 3 000 clients de votre campagne de test. Mais en augmentant la taille de votre population cible, vous augmenterez les bénéfices en conséquence.

7. Après avoir passé en revue les résultats de la simulation, sélectionnez **Utiliser pour le scoring** sous le graphique Résultats prédits pour mettre à jour les paramètres de l'onglet Evaluation en fonction du pourcentage indiqué (44 % dans le cas présent). Pour plus d'informations, voir la rubrique «Evaluation du modèle», à la page 21.
8. Si vous le souhaitez, enregistrez le projet sous *bank_response_model.str*. Dans la section suivante, vous changerez le type de modèle et créez un projet.

Modélisation de règle prédictive

Bien que les modèles prédictifs fournissent le pouvoir prédictif dont vous avez besoin, vous souhaitez peut-être obtenir un peu plus de contrôle et de connaissances sur les rouages internes du modèle dans certains cas. Pour cela, vous pouvez vous servir de la modélisation (interactive) de règle prédictive. Avec les modèles de règle prédictive, vous pouvez créer manuellement des règles pour segmenter vos données ou pour rechercher automatiquement des segments avec des concentrations élevées ou basses de la valeur du champ sélectionné. Par exemple, vous pouvez rechercher des clients qui ont répondu positivement à votre campagne puis identifier les segments avec une forte probabilité de réponse.

1. Revenez sur l'onglet Modélisation, cliquez sur **Changer de modèle** et sélectionnez **Modèle de règle prédictif**.
2. Sélectionnez les *données bancaires de réponse* pour la source de données, puis *Réponse* comme champ cible.
3. Indiquez **1** pour la réponse recherchée, pour spécifier que vous recherchez des clients ayant donné une réponse positive.
4. Sélectionnez **Démarrer la génération**.
5. Sélectionnez **Développer un modèle** pour identifier les segments ayant une probabilité de réponse plus élevée. Les trois meilleures règles qui identifient les segments présentant la probabilité de réponse la plus élevée sont créées et affichées.
6. Cliquez sur les icônes **Inclure/Exclure** pour indiquer si vous souhaitez inclure ou non les clients sélectionnés par chaque règle. Par exemple, les segments de la deuxième règle et de la troisième règle affichent des taux de réponse légèrement au-dessus de 8 %, une valeur supérieure au taux global de la campagne de test (3,7 %) mais qui reste inférieure au taux du premier segment (15,65 %). Pour connaître l'effet de l'inclusion ou de l'exclusion de ces deux règles sur votre profit global, vous pouvez exécuter la simulation de profit de ces deux façons et comparer le résultat, comme suit :
7. Après avoir défini les segments des trois règles sur **Inclure**, cliquez sur **Evaluer**.
8. Sélectionnez les types de graphique à générer suivants : **Graphique de gain** et **Graphique de profits**.

9. Sélectionnez **Simuler le profit maximum** et indiquez **3000** pour la population, **2,25** pour le coût et **60** pour le revenu. Ces mêmes caractéristiques sont utilisées lors de l'évaluation du modèle prédictif. La réponse recherchée est automatiquement complétée.
10. Cliquez sur **Exécuter** et, une fois les graphiques créés, accédez à l'onglet **Simuler**.
Le profit simulé maximum pour ce modèle est 3 178,90 euros, un montant légèrement inférieur à celui du modèle automatisé dans le cas présent. Si vous le souhaitez, vous pouvez fermer la boîte de dialogue **Evaluer**, appliquer le paramètre **Exclure** aux segments des deuxième et troisième règles, et répéter la simulation de profit pour voir l'impact sur le résultat. Les avantages réels que présente la modélisation (interactive) de règle prédictive sont le meilleur contrôle et les plus grandes connaissances qu'elle offre.
11. Si vous le souhaitez, enregistrez le projet sous *bank_predictive_rule_model.str*.

Test du modèle

Une fois votre modèle créé, vous pouvez tester une sélection d'enregistrements de client pour évaluer leurs réponses éventuelles à votre campagne.

1. Pour prévisualiser la façon dont des enregistrements spécifiques seront traités par l'évaluation du modèle, cliquez sur **Test** dans l'onglet **Modélisation**. La boîte de dialogue **Test** apparaît.
2. Dans la section **Enregistrements**, augmentez le chiffre de retour sur 100.
3. Dans la section **Sélectionner les champs à afficher**, sélectionnez *ID Client*. En outre, les champs *\$XF-Response* et *\$XFRP-Response* seront inclus, étant donné qu'ils indiquent la réponse prévue (0 pour non, 1 pour oui dans le cas présent) et le score de propension pour chaque client.

Vous pouvez ajouter, de façon facultative, des champs supplémentaires comme *Age*, *Weeks Since Last Offer* et *A un enregistrement de paiement incorrect*. Ces champs peuvent être utiles pour comprendre comment et pourquoi un client particulier est évalué de telle façon par le modèle.

Evaluation du modèle

Evaluer un modèle signifie l'appliquer à la source de données qui vous intéresse. Par exemple, après avoir généré un modèle qui prévoit les clients les plus susceptibles de répondre à une campagne, vous pouvez évaluer des prospects afin d'identifier ceux que vous voulez inclure dans votre campagne. Dans la boîte de dialogue **Evaluer**, vous avez déterminé le pourcentage d'enregistrements à évaluer afin d'optimiser le profit, et vous avez indiqué le pourcentage d'enregistrements que vous souhaitez évaluer par rapport à ce résultat.

1. Accédez à l'onglet **Evaluation**.
2. Pour **Source de données à évaluer**, assurez-vous que la source des *données bancaires de réponse* est sélectionnée.
3. Indiquez un emplacement de destination pour les résultats d'évaluation. (Si vous voulez juste prévisualiser les évaluations, vous pouvez ignorer cette étape.)
4. Sous **Champs de sortie**, assurez-vous que le champ *ID Client* est sélectionné.
5. Les paramètres **Evaluation** doivent être déjà configurés pour évaluer le même % d'enregistrements que dans la simulation de profit que vous avez exécutée lors de l'évaluation du modèle. (Si ce n'est pas le cas, sélectionnez le bouton **Utiliser pour le scoring** lors de l'affichage des résultats de la simulation ; vous pouvez y mettre les paramètres à jour.)
6. En bas de la boîte de dialogue, cliquez sur **Aperçu** pour afficher un échantillon des résultats de votre évaluation.
7. Cliquez sur **Présentation des données** pour visualiser une présentation des données que l'évaluation va utiliser. Pour plus d'informations, voir la rubrique «Présentation des données», à la page 45.
Etant donné que vous avez choisi de faire une évaluation basée sur la propension, seuls les clients ayant la propension de réponse la plus élevée sont inclus. En reliant l'*ID Client* à votre base de données client, vous pourrez générer une liste de diffusion basée sur ce résultat.

8. Cliquez sur **Evaluation** ; le modèle est appliqué aux données clients, comme le confirme le message qui s'affiche en bas de la boîte de dialogue.

Exemple de modélisation de groupement

Grâce aux modèles de groupement, vous pouvez ranger les enregistrements dans des groupes ayant des caractéristiques similaires. Vous pouvez ainsi identifier des groupes naturels dans vos données. Par exemple, vous pouvez classer les clients en fonction des caractéristiques démographiques ou du comportement d'achat.

Dans cet exemple, le fichier de données *bank_customer_data.txt*, distribué et installé avec le produit, est utilisé. Une version complète de cet exemple est également fournie dans le fichier *bank_cluster_model.str*. Contactez votre administrateur pour en savoir plus sur l'installation et pour obtenir des exemples de fichiers si nécessaire. Pour plus d'informations, voir la rubrique «Exemples de fichiers», à la page 11.

1. Revenez à la page de lancement de l'application et créez un nouveau projet IBM SPSS Modeler Advantage.
2. Sur l'onglet Données dans le panneau Sources de données du projet, ajoutez une nouvelle source de données appelée *données bancaires des clients*, comme décrit dans «Configuration des sources de données», à la page 13.
3. Dans l'onglet Modélisation, cliquez sur **Changer de modèle**, sélectionnez **Modèle de classification**, puis cliquez sur **Enregistrer**.
4. Pour la source de données, sélectionnez *données bancaires des clients*.
5. Développez la section **Paramètres facultatifs**. Pour les modèles de groupement, des paramètres facultatifs supplémentaires, **Options de cluster auto**, permettent de sélectionner un champ d'évaluation ou de définir une plage souhaitée de clusters à rechercher.

Lorsqu'un modèle de groupement est créé, un certain nombre de modèles de groupement différents sont en réalité créés en arrière-plan. Ces modèles sont ensuite comparés, puis rangés selon une mesure de qualité. Le meilleur modèle est sélectionné en vue d'une utilisation par le modèle de cluster auto. Ainsi, ces deux paramètres Options de cluster auto permettent d'exercer un certain contrôle sur la sélection de modèle. Si un champ d'évaluation est spécifié à cet endroit, le modèle qui distingue le mieux les valeurs du champ d'évaluation sera sélectionné. Et si la plage souhaitée est définie, les modèles qui identifient un certain nombre de clusters en dehors de la plage spécifiée seront supprimés.

6. Pour cet exemple, sélectionnez l'option **Définir une plage pour le nombre de classes à trouver** et utilisez les valeurs par défaut, à savoir **3** minimum et **15** maximum. Ainsi, le modèle de groupement aura un nombre de clusters adapté, ni trop, ni pas assez.
7. Les autres paramètres facultatifs sont identiques à ceux des autres types de modélisation. Décochez l'option **Nettoyer et préparer automatiquement les données pour la génération de modèle fiable** afin d'illustrer de façon plus compréhensible l'exemple de groupement automatique.
8. Réduisez le panneau **Paramètres facultatifs** et développez le panneau **Clusters**.

Le panneau Clusters comporte une section pour les clusters manuels et une section pour les clusters automatiques. Les clusters manuels vous permettent de définir des clusters en fonction de votre connaissance des données. Dans cet exemple, nous allons définir des clusters manuels pour les *clients à forte valeur* et pour les *clients jeunes et célibataires*, comme suit.

9. Cliquez sur l'icône **Créer une règle**. Appelez la règle *clients à forte valeur*, ajoutez les expressions suivantes, puis cliquez sur **OK**.

Mois en tant que client > 12
Revenus > 65 000

10. Créez une autre règle appelée *clients jeunes et célibataires* avec les expressions suivantes.

Age < 35
Etat matrimonial = C

11. Les deux nouvelles règles seront répertoriées dans la section **Clusters manuels**. Pour voir combien de clients sont identifiés par ces clusters manuels, cliquez sur **Total des enregistrements de clusters**. Ensuite, développez à nouveau les clusters manuels et consultez la colonne Nombre pour déterminer le nombre de clients identifiés pour chaque cluster. Vous constaterez que 67 clients à forte valeur et 33 clients jeunes et célibataires ont été identifiés.
12. Cliquez ensuite sur **Rechercher les clusters auto** pour rechercher automatiquement tous les autres clusters susceptibles d'être intéressants. Une fois la recherche terminée, développez la section **Clusters auto**. Vous pouvez constater que trois clusters automatiques ont été identifiés.
13. Cliquez sur **Afficher les clusters auto** pour en savoir plus. Le visualiseur des résultats des clusters automatiques s'affichera.

La page *Récapitulatif du modèle* fournit des informations de base sur l'algorithme de modèle de cluster utilisé (dans cet exemple, il s'agit de *TwoStep*), sur le nombre de clusters automatiques identifiés ainsi que sur la qualité des clusters.

D'autres graphiques sont également disponibles. Dans chaque graphique, vous pouvez survoler différentes sections pour afficher plus de détails. Vous pouvez aussi interagir avec certains des graphiques (comme le graphique d'*Importance des prédicteurs*).

Avec la vue *Clusters* (dernier graphique), vous pouvez également comparer les clusters pour rechercher des motifs intéressants. Dans une implémentation réelle, vous devrez étudier les clusters automatiques et leur donner des noms plus significatifs (en cliquant sur chaque étiquette de cluster dans la vue *Clusters*). Les noms d'étiquettes de cluster mis à jour seront alors utilisés dans tous les autres graphiques et imputés sur l'onglet Modélisation. Ensuite, vous devrez évaluer et tester le modèle, pour finalement le noter une fois que vous en êtes satisfait.

14. Si vous le souhaitez, enregistrez le projet sous *bank_cluster_model.str*.

Pour plus d'informations sur la modélisation de clusters, voir «Génération d'un modèle de groupement», à la page 63.

Exemple de modélisation d'association

La modélisation d'association est une technique importante dans l'analyse prédictive. Avec les modèles d'association, vous pouvez prévoir les éléments ayant le plus de chance d'apparaître ensemble ainsi que l'intensité de leurs relations. Par exemple, vous pouvez identifier le prochain achat le plus probable d'un client en fonction des articles dans son panier.

Dans cet exemple, le fichier de données *retail_purchase_data.txt*, distribué et installé avec le produit, est utilisé. Les données se composent d'informations d'achat concernant des articles électroniques, tels que des télévisions, des ordinateurs et des téléphones intelligents.

Une version complète de cet exemple est également fournie dans le fichier *retail_association_model.str*. Contactez votre administrateur pour en savoir plus sur l'installation des exemples de fichiers si nécessaire. Pour plus d'informations, voir la rubrique «Exemples de fichiers», à la page 11.

1. Revenez à la page de lancement de l'application et créez un nouveau projet IBM SPSS Modeler Advantage.
2. Sur l'onglet Données dans le panneau Sources de données du projet, ajoutez une nouvelle source de données appelée *données d'achat*. Sélectionnez **Fichier**, puis *retail_purchase_data.txt*.
3. Dans l'onglet Modélisation, cliquez sur **Changer de modèle**, sélectionnez **Modèle d'association**, puis cliquez sur **Enregistrer**.
4. Pour la source de données, sélectionnez *données d'achat*.
5. Pour le format de données, sélectionnez **Tabulaire**. Un ensemble de données transactionnelles afficherait uniquement les articles inclus, et tous les articles seraient présentés dans une colonne unique. Mais dans cet ensemble, chaque article est affiché dans une colonne distincte, accompagné d'un indicateur *true* (vrai) ou *false* (faux) pour indiquer s'il a été acheté ou non.

6. Cliquez sur **Générer un modèle**. Vous recevrez une erreur indiquant que la génération a échoué et que les seuils sont probablement trop hauts. Fermez l'erreur.
Lors de la recherche de règles d'association, il est généralement souhaitable de rechercher des règles applicables de façon globale, fiables et simples.
7. Développez la section Paramètres facultatifs. Sous Options de règle d'association, passez le premier seuil de 10 à 1, et le second seuil de 80 à 1. En temps normal, il serait nécessaire de tester différents seuils et de recréer le modèle à chaque fois, pour atteindre un équilibre du point de vue du nombre de règles (ni trop, ni pas assez).
8. Cliquez sur **Générer un modèle**.
Dans les résultats du modèle, chaque ligne représente une règle d'association distincte. La première règle de la liste indique que si un client achète un *téléphone intelligent*, il sera susceptible d'acheter également une *télévision grand écran*. La colonne **Couverture (%)** indique qu'environ 15 % des clients (ou environ 1 client sur 6) ont acheté un *téléphone intelligent*, et la colonne **Confiance (%)** indique que la probabilité que ces clients achètent également une *télévision grand écran* est de 30 %.
Ainsi, même avec un petit ensemble de règles (239 dans cet exemple), vous pouvez être en mesure d'utiliser ces règles dans un contexte professionnel pour prédire ou même recommander les prochains achats des clients.
9. Ensuite, cliquez sur **Test**, puis sur **Exécuter** pour voir comment les prévisions ou les recommandations sont émises.
Dans le cas présent, le modèle a consulté les règles ayant les chiffres de confiance les plus élevés, a regardé les achats déjà effectués par l'utilisateur, et à partir de ces données, a émis des recommandations sur les achats que les clients sont susceptibles de faire.
10. Cliquez sur l'icône d'aperçu en regard du premier enregistrement. Par exemple, on constate que le premier client a acheté une *télévision grand écran* et des *haut-parleurs*, mais qu'il n'a acheté aucun autre produit. Fermez l'aperçu.
La recommandation pour ce client est de lui proposer une *télévision standard*, comme indiqué dans la deuxième colonne. La valeur apparaissant dans la troisième colonne, **0,3**, indique quant à elle qu'il y a 30 % de chance qu'un client ayant déjà acheté une *télévision grand écran* et des *haut-parleurs* achètent également une *télévision standard*.
Cependant, comment savoir s'il s'agit d'une anomalie ou d'une tendance de prévision applicable de façon générale ? Notez le numéro de règle (**115** dans cet exemple) pour que nous puissions y revenir et regarder la règle plus en détail.
11. Fermez la boîte de dialogue Test.
12. Parcourez les résultats du modèle jusqu'à ce que vous parveniez au numéro de règle 115. A noter : pour les grands modèles comportant beaucoup de pages de résultats, vous pouvez également cliquer sur **Chercher une règle par ID** et entrer le numéro de règle.
13. Cliquez sur la flèche située en regard d'un en-tête quelconque de colonne de résultats et sélectionnez **Colonnes > Couverture (N)**
La colonne **Couverture (N)** s'affiche. D'après le nombre d'instances, il est évident qu'il ne s'agit pas d'une anomalie étant donné que la règle 115 se base sur 58 clients différents ayant acheté les trois produits ensemble. Cela est suffisant pour que nous puissions nous fier à la règle pour de futures recommandations.
Nous avons consulté une règle unique pour voir comment elle pouvait être utilisée pour les recommandations. Mais IBM Analytical Decision Management propose également des options intéressantes pour personnaliser le modèle afin de le faire correspondre à des fins spécifiques. Par exemple, nous souhaitons identifier les clients qui seraient intéressés pour acheter une *télévision grand écran* :
14. Cliquez sur **Appliquer les filtres**. Une boîte de dialogue s'ouvre, dans laquelle vous pouvez définir des filtres de sorte que seules les règles correspondant aux filtres soient affichées dans le modèle.
15. Sélectionnez **Autoriser le filtre**, puis sélectionnez l'article *télévision grand écran*.
16. Cliquez sur **Enregistrer** pour revenir aux résultats du modèle.

Le nombre de règles est passé de 239 à 34.

Vous souhaitez également confirmer que vous utilisez uniquement des règles basées sur un nombre raisonnable de clients. Pour cela, vous pouvez exclure les règles basées sur le comportement de moins de 10 clients :

17. Cliquez sur la flèche située en regard de l'en-tête de la colonne Couverture (N) et sélectionnez **Trier Croissant**.
18. Dans la colonne Exclure ?, sélectionnez toutes les règles ayant une valeur Couverture (N) de moins de 10 instances.
19. Notez que, même si plusieurs règles sont désormais filtrées et exclues, elles se trouvent encore dans le modèle à ce stade. Pour les supprimer de façon permanente afin qu'elles ne soient pas utilisées lors de l'évaluation du modèle, cliquez sur **Supprimer les règles exclues et filtrées**, puis sur **OK**.
Le nombre de règles applicables est réduit à 20. Toutes ces règles se basent sur au moins 10 instances et fournissent une prédiction sur l'achat d'une *télévision grand écran*.
20. Pour savoir combien de ces règles seront appliquées, cliquez sur **Test**, puis sur **Exécuter**. La *télévision grand écran* est désormais recommandée pour plusieurs autres clients avec, pour certains, un niveau de confiance plus élevé.
21. Fermez la boîte de dialogue Test.
22. Dans la section Paramètres facultatifs, développez la section Options d'évaluation pour afficher les options d'évaluation spécifiques aux modèles d'association.
Par défaut, le nombre maximum de prévisions est 3. Ce nombre se base sur les chiffres les plus élevés en termes de confiance. Vous pouvez également changer le critère de règle pour sélectionner des règles en fonction de la couverture, de la prise en charge de la règle, du montant le plus élevé de *lift* (augmentation de la probabilité d'achat d'un article spécifique) ou de la déployabilité. Vous pouvez également vouloir vérifier que vous ne proposez pas une *télévision grand écran* à des clients en ayant déjà acheté une : pour cela, sélectionnez **S'assurer qu'il n'y a pas de prédictions**.
Les options d'évaluation par défaut conviennent dans la plupart des cas. Mais dans certains cas, la flexibilité ajoutée permise par ces options peut être avantageuse.
L'étape suivante du processus serait d'utiliser l'onglet Evaluation pour évaluer le modèle. Pour plus d'informations, voir la rubrique «Evaluation (scoring) de modèles vers une table de base de données, un fichier, un serveur analytique ou un serveur IBM Cognos BI», à la page 75.
23. Si vous le souhaitez, enregistrez le projet sous *retail_association_model.str*.

Pour plus d'informations sur la modélisation d'associations, voir «Génération d'un modèle d'association», à la page 64.

Chapitre 3. Navigation dans l'espace de travail

Changement de la langue affichée

Chaque utilisateur peut facilement changer la langue de son application sans devoir redémarrer les serveurs. Par exemple, les utilisateurs de langues allemande, anglaise et japonaise peuvent tous employer la même application dans leur propre langage.

A partir du menu **Outils** en votre navigateur, accédez à **Options Internet** ou à **Options**, remplacez la valeur existante par la langue souhaitée, puis régénérez l'écran. Si IBM Analytical Decision Management n'est pas fourni dans la langue souhaitée, il s'affiche par défaut en anglais ou dans la deuxième langue sélectionnée, si celle-ci est disponible.

L'interface d'IBM Analytical Decision Management prend en charge les langues suivantes. Cette liste peut varier selon l'édition, et la documentation n'est pas disponible dans toutes les langues de l'interface utilisateur.

- Anglais
- Allemand
- Espagnol
- Français
- Italien
- Japonais
- Coréen
- Portugais (Brésilien)
- Russe
- Chinois (Chine/simplifié)
- Chinois (Taïwan/traditionnel)

Lancement d'application

La page de lancement d'application fournit des liens configurables vers les applications installées. Chaque panneau représente une application installée. Cliquez sur la liste déroulante pour sélectionner des projets, des modèles ou des règles récemment enregistrés. Vous pouvez également ajouter un ou plusieurs raccourcis personnalisés qui pointent vers des versions spécifiques marquées.

- Pour accéder à la page de lancement d'application, atteignez *http://nom_hôte:port/DM*, où *nom_hôte* est l'adresse IP ou le nom de la machine du référentiel et *port*, le numéro de port du serveur d'applications. Si vous ne connaissez pas cette information, contactez l'administrateur. Vous pouvez également ajouter dans votre navigateur un signet vers cette page de lancement ou créer un raccourci sur votre bureau électronique.

Si vous avez des problèmes d'accès à l'URL IBM SPSS Collaboration and Deployment Services ou aux URL IBM Analytical Decision Management, il est possible qu'un administrateur les ait personnalisées. Contactez votre administrateur. Pour plus d'informations, voir *IBM SPSS Collaboration and Deployment Services Repository Installation and Configuration Guide* (chapitre sur les racines de contexte d'application).

- Pour créer un nouvel objet (projet, modèle ou règle), sélectionnez **Nouveau** dans la liste déroulante pour l'application souhaitée, puis cliquez sur **Atteindre**.
- Pour ouvrir un objet (projet, modèle ou règle) enregistré, sélectionnez dans la liste déroulante la version ou le raccourci correspondant récemment utilisé. Vous pouvez également sélectionner **Parcourir** dans la liste déroulante pour sélectionner une version ou un objet différent. Pour voir rapidement l'emplacement de référentiel d'un élément de la liste, survolez cet élément dans le menu déroulant.

- Pour ajouter un nouveau raccourci personnalisé qui sera toujours disponible dans la liste déroulante, cliquez sur l'icône en forme de flèche triangulaire dans l'angle supérieur droit du panneau d'application. Vous pouvez également préciser le nombre de raccourcis à afficher.

Autres manières d'ouvrir une application

- Copiez ou entrez l'URL dans la barre d'adresses de votre navigateur, ou ajoutez l'URL à la liste de signets ou de favoris enregistrés dans votre navigateur.
- Créez un raccourci de bureau pointant vers l'URL.
- Etablissez un lien vers une URL d'application à partir du site Intranet ou du portail de votre société.

Personnalisation de la page de lancement

Les contrôles placés dans l'angle supérieur droit de la page de lancement d'application permettent d'ajouter des applications et de basculer l'affichage entre le panneau et la vue de liste. Vous pouvez également faire glisser des panneaux pour les réorganiser sur la page.

- Pour ajouter un nouveau panneau d'application, sélectionnez **Ajouter une application** dans l'angle supérieur droit de la page de lancement d'application. (Si aucune application supplémentaire n'est disponible, ce lien n'apparaît pas.)

Une liste d'applications disponibles s'affiche. En option, vous pouvez créer des raccourcis vers une ou plusieurs versions marquées.

- Pour changer la disposition des panneaux d'application sur la page de lancement, cliquez sur la barre de titre d'un panneau et maintenez le bouton de la souris pour faire glisser le panneau vers son nouvel emplacement.
- Pour supprimer une application, cliquez sur l'icône Supprimer (X rouge) dans l'angle supérieur droit du panneau d'application. Tous les raccourcis qui pointent vers cette application sont également supprimés. (L'icône Supprimer n'est pas disponible si le panneau d'application a été verrouillé sur la page par l'administrateur.)
- Les projets, les modèles et les règles non enregistrés peuvent être répertoriés dans la galerie, sur la page de lancement. Cette galerie affiche les modèles en cours de génération, ainsi que les modèles dont la génération est terminée mais qui n'ont pas encore été enregistrés. Pour plus d'informations, voir la rubrique «Galerie».

Page d'accueil

L'application fournit un flux de travail étape par étape, représenté par des icônes sur la page d'accueil. Cliquez sur une icône pour passer à l'étape correspondante.

Données. Définissez les ensembles de données que l'application doit utiliser.

Modèle. Générez des modèles à partir des données d'analyse. Évaluez-les, par exemple, à l'aide des graphiques d'importance des prédicteurs, de fréquences ou de gain afin d'obtenir des connaissances par la découverte des motifs masqués dans les données. Ensuite, employez ces modèles pour affiner vos décisions.

Evaluation. Évaluez le modèle.

Galerie

La galerie s'affiche sur la page de lancement d'application. Elle répertorie le travail non enregistré. Par exemple, lorsque vous reportez la génération d'un modèle au lendemain, ce qui vous permet de commencer le modèle avant de quitter le travail, de fermer votre navigateur, puis de revenir chercher les résultats le lendemain matin. Vous pouvez également employer la galerie pour récupérer votre travail

après la fermeture accidentelle de votre navigateur. Les éléments de la galerie peuvent inclure des projets, des règles, des modèles, des analyses de données, des évaluations et d'autres tâches à exécution longue, selon l'application.

- Si le nom d'un élément est souligné, cliquez dessus pour ouvrir l'élément correspondant. Par exemple, si vous disposez d'une analyse de données non enregistrée, cliquez sur son titre dans la galerie pour ouvrir l'onglet Données.
- Si vous n'avez pas besoin d'un élément présenté dans la galerie, cliquez sur l'icône de suppression pour le supprimer ; un message d'avertissement s'affiche pour que vous confirmiez la suppression permanente de l'élément. Pour supprimer plusieurs éléments de la galerie à la fois, cliquez sur **Vider la galerie**. La boîte de dialogue Vider la galerie s'affiche pour vous permettre de sélectionner les onglets contenant les objets à supprimer.
- Sur l'onglet Projets de la galerie, le menu déroulant **Créé dans** permet de filtrer les éléments affichés sur l'onglet pour chaque application. Par exemple, lorsque vous travaillez dans une application, vous pouvez souhaitez masquer les éléments de galerie d'autres applications.

Définition des préférences

Vous pouvez ajuster un certain nombre de préférences pour personnaliser IBM Analytical Decision Management selon vos besoins. Principalement, cette personnalisation consiste à définir vos propres options par défaut pour vous épargner de sélectionner plusieurs fois les mêmes actions lorsque vous créez, testez et enregistrez votre travail. Les préférences que vous définissez vous sont propres et s'appliquent à toutes les applications que vous utilisez.

Pour accéder aux paramètres des préférences utilisateur, sur la barre d'outils de toute application, cliquez sur l'icône **Définir les préférences de l'utilisateur**.



Figure 9. Icône Définir les préférences utilisateur

Vous pouvez préciser vos paramètres préférés pour chaque secteur, comme indiqué dans les sections qui suivent.

Paramètres généraux

Nombre par défaut d'objets récemment ouverts dans les raccourcis d'application. Sélectionnez le nombre d'éléments qui doivent être disponibles dans la liste déroulante affichée sur la page de lancement d'application.

Format de date. Sélectionnez le format à utiliser pour afficher ou enregistrer des dates.

Format d'heure. Sélectionnez le format à utiliser pour afficher ou enregistrer l'heure.

Devise. Sélectionnez la devise par défaut pour chaque nouveau projet. Cette valeur ne remplace pas les données existantes entrées dans une devise différente.

Décimales. Sélectionnez le nombre de positions décimales à afficher pour les nombres.

Nombre de lignes à afficher lors de l'aperçu. Sélectionnez le nombre maximal d'enregistrements à afficher dans les aperçus (par exemple, pour vérifier une source de données lors de l'importation).

Nombre de résultats par page (tables paginées). Sélectionnez le nombre maximal de résultats à afficher sur chaque page lors de l'affichage de tables.

Nombre maximal d'éléments affichés dans un graphique. Parfois, la représentation des résultats dans les graphiques peut être incorrecte lorsque de nombreux éléments doivent être affichés. Indiquez le nombre maximal d'éléments à afficher dans les graphiques. Ce paramètre force l'affichage d'une table plutôt que d'un graphique lorsque le nombre d'éléments dépasse le seuil indiqué. Il ne s'applique qu'aux champs catégoriels des présentations de données et au graphique du rapport récapitulatif d'évaluation des modèles dotés d'une cible catégorielle.

Nombre de segments de données (mosaïques) à appliquer aux graphiques et aux tables. Sélectionnez le nombre maximal de segments de données à utiliser dans toute l'application pour les graphiques et les tables, selon le cas.

Paramètres de modèle

Génération du type de modèle par défaut. Indiquez l'option par défaut souhaitée : l'application doit générer le meilleur modèle automatiquement à partir des données disponibles (modèle prédictif) ; un modèle doit identifier les segments qui présentent une probabilité plus ou moins élevée par rapport à un résultat spécifique (modèle de règle prédictif) ; un modèle doit catégoriser des enregistrements dans un certain nombre de clusters (modèle de groupement) ; un modèle doit prévoir les éléments qui sont davantage susceptibles d'apparaître ensemble ainsi que la puissance de leurs interrelations (modèle d'association). Seuls les types de modèle configurés pour apparaître dans l'application sont disponibles dans le menu déroulant.

Partitionner automatiquement les données pour activer le scoring du modèle dans la source de données de génération pour le scoring et le test. La sélection de cette option divise les données en sous-ensembles ou échantillons distincts à des fins d'apprentissage et de test du modèle. En générant le modèle sur un sous-ensemble et en le testant sur un autre, vous pouvez avoir une idée de la façon dont il va se généraliser à d'autres ensembles de données. Vous pouvez également indiquer un pourcentage pour la *Taille de la partition d'apprentissage (%)* et la *Taille de la partition de test*. Ces options sont disponibles pour tous les types de modèle excepté les modèles d'association.

Nettoyer et préparer automatiquement les données pour la génération de modèle fiable. Sélectionnez cette option pour laisser l'application prendre en charge la plupart des tâches de préparation de données. Par exemple, résoudre les problèmes de données, examiner les données les moins utiles ou dériver de nouveaux attributs.

Vous pouvez indiquer si vous souhaitez accorder une importance égale à la vitesse et la précision lors de l'opération de nettoyage, faire prévaloir la précision, ou accorder la priorité à la vitesse (par exemple, pour travailler avec de grands ensembles de données ou pour rechercher un résultat rapide).

Remarque : Les paramètres de préparation des données de mémoire automatique s'appliquent uniquement lors de la création d'un nouveau modèle et la mise au rebut n'est pas effectué.

Limiter le temps de génération d'un seul modèle. Utilisez cette option pour indiquer le temps par défaut de génération de chaque modèle. Cette option peut être particulièrement utile lorsque vous traitez de grands ensembles de données et peut avoir un impact important sur le temps nécessaire pour exécuter la génération de modèle.

Paramètres de test et de simulation

Nombre par défaut d'enregistrements testés. Indiquez le nombre par défaut d'enregistrements à inclure lors de l'exécution d'un test.

Date de test/simulation par défaut. Indiquez si la date de simulation par défaut est définie sur la date du jour ou sur une date spécifique.

Ouverture et enregistrement de projets, de modèles et de règles

Vous pouvez ouvrir, supprimer ou enregistrer des fichiers dans IBM SPSS Collaboration and Deployment Services Repository, ou les télécharger dans le système de fichiers local. Vous pouvez également télécharger des fichiers de flux IBM SPSS Modeler à partir du système de fichiers local et les enregistrer dans IBM SPSS Collaboration and Deployment Services Repository. IBM SPSS Collaboration and Deployment Services Repository offre d'importants avantages, dont le support de plusieurs versions, le marquage avec des libellés et la sécurité. Selon la configuration locale, certaines options peuvent ne pas être prises en charge.



Crée un nouveau modèle.



Enregistre le projet en cours dans son dernier emplacement connu dans IBM SPSS Collaboration and Deployment Services Repository. Vous pouvez également appliquer un nouveau libellé lors de l'enregistrement.



Enregistre le projet en cours en lui donnant un nouveau libellé ou sous un nouveau nom. Le cas échéant, vous pouvez enregistrer l'objet dans IBM SPSS Collaboration and Deployment Services Repository ou le télécharger dans le système de fichiers local.



Ouvre un projet à partir du IBM SPSS Collaboration and Deployment Services Repository ou du système de fichiers local selon le cas. Vous pouvez également, le cas échéant, télécharger un projet à partir du système de fichiers local. L'accès au système de fichiers n'est pas pris en charge pour toutes les applications. Si vous téléchargez un flux IBM SPSS Modeler, vous pouvez ouvrir le fichier ou de l'enregistrer dans le référentiel.

Ouverture depuis et enregistrement dans le référentiel

IBM SPSS Collaboration and Deployment Services Repository offre d'importants avantages, dont le support de plusieurs versions, le marquage avec des libellés et la sécurité. Chaque fois que vous enregistrez ou créez une nouvelle version, vous pouvez revenir à une version antérieure à tout moment. Vous pouvez également gérer les versions à l'aide de libellés et appliquer des paramètres de sécurité pour déterminer l'utilisateur autorisé à consulter, à visualiser ou à supprimer un objet.

Dans le volet gauche, l'arborescence affiche les dossiers d'IBM SPSS Collaboration and Deployment Services Repository que vous êtes autorisé à visualiser. Naviguez dans l'arborescence des dossiers ou recherchez l'élément voulu. Vous pouvez également filtrer la liste de fichiers à l'aide de libellés disponibles dans le volet gauche inférieur. Par exemple, vous pouvez cliquer sur un libellé de *production* pour n'afficher que les fichiers marqués avec ce libellé.

Pour plus de détails sur un fichier, le cas échéant, cliquez sur l'icône d'aperçu ou d'information située en regard.

Pour supprimer un fichier, cliquez l'icône de suppression située en regard de celui-ci. Cette fonction est particulièrement utile à ceux qui n'ont pas accès au logiciel client IBM SPSS Collaboration and Deployment Services Deployment Manager.

Dossier. Affiche les dossiers disponibles et permet de naviguer dans la hiérarchie. Sélectionnez le dossier à enregistrer. Vous pouvez également créer un nouveau dossier ou renommer un dossier existant à l'aide des icônes au-dessus de la liste de dossiers.

Filtrer par étiquette. Limite l'affichage aux objets dotés du libellé sélectionné. Lors de la recherche de dossiers, seuls les objets portant ce libellé s'affichent.

Lors de l'exportation de règles utilisées dans une application, une liste déroulante **Enregistrer sous** ou **Exporter en tant que** est également disponible. Sélectionnez **Jeu de règles unique** ou **Règles individuelles** dans la liste déroulante pour indiquer le mode d'enregistrement de la règle. Si seules les règles référencées sont disponibles, la liste déroulante est désactivée, et la règle est enregistrée en un jeu de règles unique (les règles référencées ne peuvent pas être enregistrées individuellement).

Si vous sélectionnez l'enregistrement en tant que règles individuelles, une table s'affiche pour vous permettre d'entrer les détails de chaque règle à enregistrer.

Nom. Entrez un nom unique pour l'objet que vous enregistrez.

Mots-clés. Entrez des mots clés pour identifier rapidement l'objet enregistré lors de sa recherche. Par exemple, *Ventes* ou *1-2009*.

Description. Entrez une description longue de l'objet enregistré. Par exemple, *Analyse des ventes au détail ; 1er trimestre 2009*.

Étiquette(s) Si vous risquez de disposer de plusieurs versions d'un même objet enregistré, vous pouvez identifier chaque version avec un libellé (ou une étiquette) unique. Sélectionnez un libellé existant ou créez-en un.

Utilisation des libellés

Les libellés (ou étiquettes) permettent d'identifier une version spécifique d'un projet, d'un modèle ou d'une règle. Lorsque vous partagez une règle entre plusieurs projets, vous pouvez préciser la version de cette règle à utiliser (par exemple, la version marquée *production*). Si une version plus récente de la règle est créée, celle utilisée dans votre projet ne change pas sauf si le libellé *production* est explicitement déplacé vers la nouvelle version.

Lors du référencement d'objets partagés, il est fortement recommandé d'utiliser une version marquée, en particulier dans un environnement de production, car les changements apportés à la version la plus récente peuvent avoir des conséquences imprévues. Dans certains cas, par exemple lorsque vous faites référence à une règle à partir d'une autre règle, une version marquée est obligatoire. Pour plus d'informations, voir la rubrique «Exportation et réutilisation de règles», à la page 56.

Si aucune version n'est indiquée, la version la plus récente est utilisée, ce qui signifie que les changements ultérieurs apportés à la règle sont automatiquement pris en compte. (*LATEST* est un libellé spécial automatiquement déployé à chaque nouvelle version créée.)

Chaque fois que vous enregistrez un nouvel objet ou une nouvelle version, vous pouvez sélectionner les libellés applicables. Pour plus d'informations, voir la rubrique «Ouverture et enregistrement de projets, de modèles et de règles», à la page 31.

Restrictions sur les noms de libellé. Lorsque vous définissez des libellés utilisables avec IBM Analytical Decision Management, vous ne pouvez pas inclure les caractères **?** **#** **&** et **:** (deux-points) dans les noms de libellé car ils peuvent empêcher l'application d'interpréter correctement les références aux objets qui utilisent ces libellés. Les noms de libellé suivants, par exemple, ne sont pas valides :

Tester & Déployer

DEPLOYER:PHASE1

Verrouillage des projets

Lorsque vous ouvrez un projet, un modèle ou une règle, une case à cocher Verrou situé en haut à droite vous permet de verrouiller l'objet dans IBM SPSS Collaboration and Deployment Services Repository, si bien que les autres utilisateurs ne peuvent pas le modifier.

- Lorsqu'un projet est verrouillé, les autres utilisateurs peuvent l'ouvrir pour le visualiser, mais ils ne peuvent enregistrer aucun changement.
- Lorsque vous fermez un projet verrouillé (par exemple, en fermant le navigateur), le verrou est libéré pour que d'autres utilisateurs puissent éditer l'objet.
- Si vous sélectionnez de déverrouiller un projet sans le fermer (en décochant la case Verrou), vous serez invité à enregistrer les éventuels changements. Les changements non enregistrés seront perdus après le déverrouillage de l'objet.
- Après avoir déverrouillé un projet, vous pouvez continuer à le modifier, sauf si un utilisateur le verrouille pendant ce temps. Dans ce cas, vous ne pourrez plus enregistrer vos modifications.

Propriétés

La boîte de dialogue Propriétés inclut un onglet **Projet** et un onglet de **référentiel**. Pour afficher la boîte de dialogue Propriétés, cliquez sur l'icône Propriétés dans n'importe quelle application.



Figure 10. icône Propriétés

Sur l'onglet **Projet**, vous pouvez indiquer le nombre maximal de membres autorisés pour les champs *nominaux*. Cette fonction est utilisée sur l'onglet Données. Sélectionnez l'option correspondant au **nombre maximal de membres autorisés pour les champs nominaux**. La valeur par défaut est **250**. Cette fonction est utile par exemple dans les situations suivantes :

1. Un utilisateur crée une nouvelle source de données et exécute une analyse de données.
2. Un certain nombre de champs avec des informations utiles sur les catégories apparaissent dans le modèle de données du projet comme *Sans type* car le nombre de valeurs distinctes dans les données dépasse la limite de taille maximale par défaut définie dans le dialogue Propriétés.
3. L'utilisateur clique sur l'icône Propriétés et augmente le **nombre maximal de membres autorisés pour les champs nominaux** pour qu'il soit approprié à ses données.
4. De nouveau sur l'onglet Données, l'utilisateur clique sur **Actualiser l'analyse de données** mais les champs restent *Sans type* car l'action d'actualisation ne modifie pas la mesure, pour conserver toutes modifications apportées par l'utilisateur. En d'autres termes, l'actualisation conserve la mesure existante pour un champ et recalcule uniquement les valeurs parce que toutes les modifications apportées par l'utilisateur dans le modèle de données doivent être conservées.
5. Pour utiliser les valeurs dans le modèle de données du projet, l'utilisateur doit modifier manuellement la mesure de chaque champ de *Sans type* à *Catégoriel* avant d'exécuter l'actualisation. Vous remarquerez que, si de nombreux champs sont concernés, il peut être plus rapide de recréer la source de données.

Le paramètre du nombre maximal de membres autorisés pour les champs nominaux ne s'applique pas aux sources de données du type de fichier *.sav*. De nombreux fichiers *.sav* incluent des fichiers et valeurs personnalisés. Ils ne sont donc pas concernés par ce paramètre. L'utilisateur doit modifier manuellement les champs.

Sur l'onglet du **référentiel**, vous pouvez ajouter des mots clés et une description pour identifier un projet, un modèle ou un ensemble de règles lorsque vous les rechercherez à une date ultérieure :

- **Mots-clés.** Entrez des mots clés pour identifier rapidement l'objet enregistré lors de sa recherche. Par exemple, *Ventes* ou *Q1-2013*.

- **Description.** Entrez une description longue de l'objet enregistré. Par exemple, *Analyse des ventes au détail ; premier trimestre 2013.*

Chapitre 4. Gestion des sources de données

L'onglet Données permet de définir des sources de données pour l'analyse, la simulation, les tests, l'évaluation ou autre. Les ensembles de données peuvent être configurés par l'administrateur. Vous pouvez également ajouter les vôtres.

L'onglet inclut un onglet **Source** et un onglet **Dérivé**. L'onglet se divise en deux zones principales :

- **Modèle de données du projet.** Définit les zones requises par l'application. Tous les autres ensembles de données sont mappés par rapport à cette source. La section **Champs de source de données** répertorie les champs d'entrée et les types de données qu'utilise votre projet. Au besoin, utilisez l'onglet **Dérivé** afin d'ajouter des expressions ou des données de sortie de modèle de différentes sources de données pour étendre le modèle de données du projet.
- **Sources de données.** La section **Sources de données du projet** répertorie les sources de données qui ont été enregistrées en vue de leur utilisation avec le projet, le modèle ou la règle en cours. La section **Mes sources de données** répertorie les sources de données que vous avez définies ou enregistrées. Les ensembles de données de cette liste peuvent être copiés dans ou depuis différents projets, puis promus pour être utilisés en tant que modèle de données du projet, ce qui permet de partager les données entre les applications.

Modèle de données du projet

Le modèle de données du projet sert de canevas répertoriant les zones d'entrée et les types de données que le projet va utiliser.

Source de données. Affiche la source de données sélectionnée pour l'application.

ID enregistrement. Au besoin, vous pouvez sélectionner un champ à utiliser comme ID d'*entité* pour activer le processus de renvoi SQL push-back. Cette sélection permet d'améliorer les performances lorsque les données proviennent d'une base de données et que le champ sélectionné est un entier. Dans IBM SPSS Modeler Advantage, le champ sélectionné ici sera par défaut sélectionné comme sortie d'une évaluation de lot et désélectionné par défaut en tant qu'entrée d'une génération de modèle. Ce champ est facultatif.

Opérationnel. Cochez cette case pour indiquer que les données de la zone sont disponibles pour une utilisation dans les règles et les expressions, ainsi que dans les prévisions. Décochez la case si le champ doit être utilisé comme champ cible pour l'analyse. Par exemple, si vous exécutez une requête pour voir comment les clients vont réagir à une campagne de publipostage, le champ des résultats sera analytique car vous ne connaissez pas à l'avance les valeurs des réponses.

Nom de la zone. Affiche le nom de chaque zone, énumérée dans la source de données, avec une icône qui identifie son type de stockage.

Mesure. Affiche le type de mesure de chaque zone dans la source de données. Si vous changez l'un des types de mesure, vous devez actualiser les données d'analyse pour garantir la compatibilité des données. Pour plus d'informations, voir la rubrique «Niveaux de mesure», à la page 46.

Valeurs. Répertorie les valeurs de chaque zone dans la source de données. Par exemple, les valeurs maximale et minimale d'une plage. Pour les zones d'indicateur, placez le pointeur de la souris sur la valeur pour afficher une infobulle qui indique les valeurs vraies ("true") et fausses ("false") définies pour la zone. Vous pouvez également cliquer sur une valeur pour l'éditer.

En général, le modèle de données du projet définit les entrées comme un ensemble de zones provenant d'une unique table logique. C'est toujours le cas pour IBM SPSS Modeler Advantage. En revanche, pour d'autres applications, le modèle de données du projet peut consister en une table principale accompagnée de 0 ou plusieurs tables secondaires. Ces tables secondaires sont définies par les sources de données secondaires associées à la source de données principale.

Sources de données du projet

Les sources de données du projet peuvent inclure celles qui sont prédéfinies par l'administrateur ou ajoutées par les utilisateurs. Facultativement, l'administrateur peut avoir verrouillé une ou plusieurs sources de données pour empêcher les utilisateurs de les modifier ou de les retirer, ou verrouillé toutes les options de données pour que les utilisateurs ne puissent pas créer de nouvelles sources de données.

Si les entrées d'une source de données ne correspondent pas directement à celles du modèle de données du projet, vous pouvez mapper la première au second et corriger les différences. Par exemple, si le modèle de données du projet nécessite un champ nommé *achat* avec les valeurs *Oui* et *Non* (indicateur de niveau de mesure), toute source de données utilisée doit disposer d'une zone comparable pouvant être mappée en conséquence.

Nom. Affiche le nom de la source de données et une icône qui identifie son type de fichier.

Nombre d'enregistrements. Cliquez sur l'icône dans cette colonne pour afficher le nombre d'enregistrements de toute source de données répertoriée dans la table. Ce nombre s'affiche en regard de l'icône.

Aperçu. Cliquez sur l'icône dans cette colonne pour prévisualiser un échantillon des données contenues dans la source. Pour plus d'informations, voir «Aperçu des données», à la page 45.

Généralités. Cliquez sur l'icône dans cette colonne pour obtenir une présentation de la source de données. Pour plus d'informations, voir «Présentation des données», à la page 45.

Compatible. Affiche une note informant que la source de données est utilisée comme modèle de données du projet ou affiche une boule verte, orange ou rouge pour indiquer le niveau de compatibilité de la source de données avec le modèle de données du projet.

- Une boule verte indique que la source de données est opérationnellement compatible avec la source de données du modèle de données du projet. Une source de données opérationnellement compatible est une source de données qui inclut toutes les zones opérationnelles du modèle de données du projet, mais qui peut comporter des zones supplémentaires. Cette source de données convient pour les règles, l'évaluation (scoring), la simulation et les opérations de test.
- Une boule orange indique que la source de données comporte au moins une zone compatible avec le modèle de données du projet, de même nom et type. Cette source de données peut également comporter d'autres zones et convient pour la génération et l'évaluation de modèles.
- Une boule rouge indique que la source de données est incompatible avec le modèle de données du projet et que ses zones doivent être mappées pour être utilisées dans l'application. Une source de données incompatible comporte au moins une zone dont le type n'est pas compatible avec le type équivalent du modèle de données du projet.

Mapper les zones. Cette option permet de comparer les zones de source de données avec celles du modèle de données du projet ; vous pouvez mapper ou démapper des zones compatibles pour les faire correspondre avec celles exigées par le modèle de données du projet. Pour plus d'informations, voir «Mappage de champs», à la page 47.

Copier. Copie la source de données dans la section **Mes sources de données**.

Mes sources de données

Les sources de données figurant sur cette liste sont enregistrées avec votre compte utilisateur. Par conséquent, elles sont disponibles chaque fois que vous vous connectez et peuvent être copiées dans tout projet, tout modèle ou toute règle que vous ouvrez (si l'administrateur vous a octroyé les droits appropriés). Les zones de cette partie de l'onglet fonctionnent de la même façon que celles de la section **Sources de données du projet**, bien qu'il n'existe pas de colonne **Compatibilité**.

La colonne **Copier** permet de copier la source de données dans la section **Sources de données du projet**.

Utilisation des sources de données

- Pour ajouter une nouvelle source de données à la liste **Sources de données du projet** ou **Mes sources de données**, sélectionnez **Ajouter une source de données**. Pour plus d'informations, voir «Création d'une source de données».
- Pour changer le niveau de mesure d'une zone de la source de données dans le **Modèle de données du projet**, sélectionnez le niveau qui convient. Pour plus d'informations, voir «Niveaux de mesure», à la page 46.
- Pour copier des sources de données vers ou depuis la liste **Sources de données du projet**, cliquez sur la flèche appropriée dans la colonne **Copier**.
- Pour mapper les noms de zone d'une source de données avec le modèle de données du projet, cliquez sur le lien approprié sous la colonne **Compatible**. (Ce lien ne s'affiche plus une fois que les noms de champ ont été mappés.) Pour plus d'informations, voir «Mappage de champs», à la page 47.
- Pour prévisualiser une source de données, cliquez sur l'icône **Aperçu**. Pour plus d'informations, voir «Aperçu des données», à la page 45.
- Pour ajouter les champs supplémentaires (expressions, champs utilisant des règles de segment ou sortie de modèle de différentes sources de données) afin d'étendre votre modèle de données du projet, cliquez sur l'onglet **Dérivé**. Pour plus d'informations, voir «Calcul des champs», à la page 48.

Création d'une source de données

1. Pour accéder à l'éditeur de source de données, sélectionnez **Ajouter une source de données** sous la section "Sources de données de projet" ou "Mes sources de données".
2. Entrez un nom pour la source de données.
3. Sélectionnez le type de source. Remarque : la liste des bases de données disponibles fait partie de celles disponibles sur la machine IBM SPSS Modeler.
 - **Fichier**. De nombreux types de fichier sont pris en charge, dont le format texte (*.txt, *.csv, etc.), Excel, les formats *.sav et *.zsav utilisés par des fichiers IBM SPSS Statistics et SAS. Pour plus d'informations, voir la rubrique «Source Fichier», à la page 38.
 - **Base de données**. Toute base de données acceptant ODBC, dont Microsoft SQL Server, DB2, Oracle. Pour plus d'informations, voir la rubrique «Source Base de données», à la page 41.
 - **Vue de données**. Crée une connexion à utiliser avec des règles IBM Operational Decision Management. Les administrateurs peuvent restreindre l'utilisation de règles complexes qui emploient un modèle de données hiérarchique car les règles de ce type peuvent dégrader les performances. Pour plus d'informations, voir la rubrique «Source Vue de données», à la page 42.
 - **Serveur IBM Cognos BI**. Toute base de données acceptant le serveur IBM Cognos BI. Pour plus d'informations, voir la rubrique «Source IBM Cognos BI», à la page 42.
 - **Analytic Server**. Vous ne pouvez pas sélectionner ce type de source lorsque vous créez une nouvelle source de données. IBM SPSS Modeler 16 a introduit IBM SPSS Analytic Server, qui permet d'utiliser HDFS (Hadoop Distributed File System) comme source de données. Les utilisateurs d'IBM SPSS Modeler peuvent créer des flux à l'aide du noeud source HDFS Analytic Server et générer des modèles d'après le système de volumes massifs de données (big data)

Hadoop. Les utilisateurs d'IBM Analytical Decision Management peuvent ouvrir ces flux de modèles dans IBM SPSS Modeler Advantage pour les visualiser (ils ne peuvent pas créer de projets à l'aide de sources de données Analytic Server).

4. Indiquez les options appropriées pour le type de source.

Vous pouvez cliquer sur l'icône **Nombre d'enregistrements** pour afficher le nombre d'enregistrements contenus dans la source de données en cours, ou cliquer sur **Aperçu** pour afficher les dix premiers enregistrements de la source de données.

Source Fichier

L'application prend en charge plusieurs types de fichier de données. Lorsque vous entrez ou recherchez un nom de fichier, l'application détecte automatiquement le type de fichier et développe l'éditeur de source de données pour afficher d'autres champs relatifs à ce type de fichier. Si le type de fichier est incorrect ou nécessite une version spécifique (par exemple, Microsoft Excel 1997-2003), vous pouvez le remplacer par le type de fichier requis.

Vous pouvez cliquer sur **Parcourir** pour sélectionner un fichier dans le référentiel, ou sur **Télécharger le fichier local** pour sélectionner un fichier dans votre système de fichiers local (si activé).

Si vous téléchargez un fichier local, le téléchargement s'effectue dans IBM SPSS Modeler Server. Dans la boîte de dialogue Télécharger le fichier, recherchez et sélectionnez le fichier local à télécharger, puis la destination de téléchargement IBM SPSS Modeler Server. Lorsque vous sélectionnez la destination, vous pouvez changer le nom de fichier si vous le souhaitez.

La fonction **Télécharger le fichier local** est désactivée par défaut. Les administrateurs peuvent utiliser l'outil basé sur navigateur, IBM SPSS Collaboration and Deployment Services Deployment Manager, pour activer ou désactiver cette fonction, ainsi que pour augmenter ou réduire la taille de fichier maximale autorisée.

Fichiers de données à base de texte

Lorsque vous sélectionnez une source de données à base de texte, vous êtes invité à entrer d'autres détails.

La première ligne a des noms de colonne. Sélectionnez cette option si le nom de chaque colonne est inclus comme ligne d'en-tête dans la source de données.

Codage. Indique la méthode de codage de texte employée. Vous pouvez sélectionner la valeur par défaut du système ou UTF-8.

Symbole décimal. Indique le mode de représentation des décimales dans les données.

- **Point (.)**. Le point est utilisé comme séparateur décimal.
- **Virgule (,)**. La virgule est utilisée comme séparateur décimal.

Délimiteurs. A l'aide des cases à cocher répertoriées pour cette commande, vous pouvez spécifier quels caractères (comme la virgule) marquent les limites des champs dans le fichier. Vous pouvez indiquer plusieurs délimiteurs(, | par exemple) pour les enregistrements qui font appel à des délimiteurs multiples. Le séparateur par défaut est la virgule.

Remarque : si la virgule est également définie en tant que séparateur décimal, les paramètres par défaut fournis ne fonctionnent pas. Si la virgule sert à la fois de séparateur de champs et de séparateur décimal, sélectionnez **Autre** dans la liste Séparateurs. Ensuite, ajoutez manuellement une virgule dans le champ d'entrée.

Sélectionnez **Autoriser plusieurs délimiteurs vides** pour considérer plusieurs délimiteurs vides adjacents comme un délimiteur unique. Par exemple, une séquence constituée d'une valeur de données suivie de quatre espaces, puis d'une autre valeur de données, sera considérée comme séquence à deux champs, et non comme une séquence à cinq champs.

Options avancées

Caractères de commentaires fin de ligne. Spécifiez quels caractères (comme # ou !) indiquent des annotations dans les données. Lorsque l'un de ces caractères apparaît dans le fichier, toutes les données situées entre ce caractère et le caractère de nouvelle ligne suivant (non inclus) sont ignorées.

Indiquer les champs d'entrée. Indiquez le nombre de champs d'entrée à extraire de chaque enregistrement.

Spécifier le format de données Pour les sources de données **Fichier** qui sont de type **Champs de longueur variable en texte brut**, vous pouvez utiliser cette section pour définir le type de stockage d'entrée et le format de chaque champ afin de vous assurer que les valeurs sont lues correctement. Cette procédure est similaire à la fonctionnalité disponible dans le client IBM SPSS Modeler. L'option **Remplacer** indique si la valeur par défaut est remplacée. Si vous sélectionnez **Remplacer**, cela active les commandes de **Stockage** et de **Format d'entrée**. Le fait de désélectionner **Remplacer** redéfinit les valeurs d'origine par défaut. **Format d'entrée** applique uniquement les types de stockage réel, date, heure et horodatage.

Ignorer les caractères des en-têtes. Indiquez le nombre de caractères à ignorer au début du premier enregistrement.

Lignes à analyser pour le noeud Typer. Indiquez le nombre de lignes à traiter pour les types de données indiqués.

Supprimer les espaces de début et de fin. Cochez cette case pour que les espaces situés en début et en fin des chaînes soient supprimés lors de l'importation. Vous pouvez sélectionner la suppression des espaces à gauche, à droite, des deux côtés ou aucun.

Caractères non valides. Sélectionnez **Supprimer** pour supprimer les caractères non valides de la source de données. Sélectionnez **Remplacer par** pour remplacer les caractères non valides par le symbole indiqué (un caractère uniquement). Les caractères non valides sont des caractères nuls (0) ou des caractères qui n'existent pas dans la méthode de codage spécifiée.

Guillemets. A l'aide des listes déroulantes, vous pouvez indiquer la façon dont les guillemets simples et doubles sont traités lors de l'importation. Vous pouvez choisir l'option **Supprimer** (supprime tous les guillemets), **Inclure comme texte** (inclut les guillemets dans la valeur de la zone) ou **Apparier et supprimer** (supprime des paires de guillemets). Si un guillemet n'est pas apparié, un message d'erreur apparaît. Si vous sélectionnez **Supprimer** ou **Apparier et supprimer**, la valeur de la zone (sans les guillemets) est stockée sous forme de chaîne.

Remarque : Lorsque vous utilisez **Apparier et supprimer**, les espaces sont conservés. Lorsque vous utilisez **Supprimer**, espaces en début et en fin à l'intérieur et à l'extérieur des guillemets sont supprimés (par exemple, ' " ab c" ', "d ef " ', " gh i " ' donnent 'ab c, d ef, gh i'). Lorsque vous utilisez **Inclure comme texte**, les guillemets sont considérés comme des caractères normaux ; les espaces de début et de fin sont donc supprimés.

Indiquer les champs d'entrée. Pour plus d'informations, voir la rubrique «Sélection de champs d'entrée», à la page 45.

Fichiers de données Excel

Lorsque vous sélectionnez une source de données Excel, vous êtes invité à entrer d'autres détails :

La première ligne a des noms de colonne. Sélectionnez cette option si le nom de chaque colonne est inclus comme ligne d'en-tête dans la source de données.

Plage nommée. Permet de sélectionner une plage nommée de cellules, définie dans la feuille de calcul Excel. Si vous utilisez une plage nommée, les autres paramètres de feuille de calcul et de plage de données ne s'appliquent plus et sont désactivés en conséquence.

Choisir une feuille de calcul. Indique la feuille de calcul à importer, par son nom ou par son index.

- **Par nom.** Sélectionnez le nom de la feuille de calcul à importer.
- **Par index** Définit la valeur d'index de la feuille de calcul à importer, 0 désignant la première feuille de calcul, 1, la deuxième et ainsi de suite.

Intervalle sur feuille de calcul. Vous pouvez importer des données en partant de la première ligne renseignée ou en indiquant un intervalle de cellules explicite.

- **L'intervalle commence à la première ligne non renseignée.** Repère la première cellule renseignée et l'utilise comme angle supérieur gauche de l'intervalle de données.
- **Intervalle de cellules explicite.** Vous permet de spécifier un intervalle explicite par ligne et par colonne. Par exemple, pour spécifier l'intervalle Excel A1:D5, vous pouvez entrer A1 dans le premier champ et D5 dans le second (ou bien, R1C1 et R5C4). Toutes les lignes de l'intervalle indiqué sont renvoyées, y compris les lignes vides.

Sur les lignes vides. Si vous rencontrez une ligne vide, vous pouvez l'ignorer ou sélectionner **Retourner les lignes vides** pour continuer la lecture de toutes les données jusqu'à la fin de la feuille de travail, y compris les lignes vides.

Indiquer les champs d'entrée. Pour plus d'informations, voir la rubrique «Sélection de champs d'entrée», à la page 45.

Fichiers de données IBM SPSS Statistics

Quand vous sélectionnez une source de données IBM SPSS Statistics (fichier *.sav* ou *.zsav*), vous devez, sur invite, entrer des détails supplémentaires. Si le fichier est protégé par mot de passe, vous devez également, sur invite, entrer le mot de passe.

Noms des variables. Sélectionnez une méthode de gestion des noms de variable et des libellés, applicable lors de l'importation à partir d'un fichier IBM SPSS Statistics *.sav* ou *.zsav*.

- **Lire les noms et les libellés.** Sélectionnez cette option pour lire les noms de variable et les libellés ; il s'agit de l'option par défaut. Les libellés peuvent être affichés dans les graphiques, dans les explorateurs de modèles et dans d'autres types de sortie.
- **Lire les libellés sous forme de nom.** Sélectionnez cette option pour lire les libellés de variable dans le fichier *.sav* IBM SPSS Statistics plutôt que les noms de champ courts, puis utilisez ces libellés comme noms de variable.

Valeurs. Sélectionnez une méthode de gestion des valeurs et des libellés lors de l'importation à partir d'un fichier IBM SPSS Statistics *.sav* ou *.zsav*.

- **Lire les données et les libellés.** Sélectionnez cette option pour lire les valeurs réelles et les libellés de valeur ; il s'agit de l'option par défaut.
- **Lire les libellés sous forme de données.** Indiquez si vous souhaitez utiliser les libellés de valeur à partir du fichier *.sav* ou *.zsav* au lieu des codes numériques ou symboliques utilisés pour représenter les valeurs. Par exemple, si vous sélectionnez cette option pour les données dont le champ indiquant le

genre a pour valeur 1 et 2 (représentant respectivement *masculin* et *féminin*), le champ sera converti en chaîne, et importera *masculin* et *féminin* comme valeurs réelles.

Il est important de prendre en compte les valeurs manquantes dans vos données IBM SPSS Statistics avant de choisir cette option. Par exemple, si un champ numérique utilise des libellés uniquement pour les valeurs manquantes (0 = *Pas de réponse*, 99 = *Inconnu*) et que vous sélectionnez l'option ci-dessus, seuls les libellés de valeurs *Pas de réponse* et *Inconnu* sont importés, et le champ est converti en chaîne. Dans ce cas, vous devez importer les valeurs elles-mêmes.

Indiquer les champs d'entrée. Pour plus d'informations, voir la rubrique «Sélection de champs d'entrée», à la page 45.

Source Base de données

Lorsque vous sélectionnez une source Base de données, vous êtes invité à entrer d'autres détails :

Base de données. Sélectionnez la base de données à utiliser.

Connexion Sélectionnez la connexion à la base de données en utilisant les données d'identification de la base de données ou les données d'identification stockées dans IBM SPSS Collaboration and Deployment Services Repository. Puis, entrez votre nom et votre mot de passe de connexion à la base de données ou sélectionnez des données d'identification dans le référentiel et cliquez sur **Connecter** pour vérifier que vous avez accès à la base de données.

Notez que le nom et le mot de passe de connexion à la base de données ne sont pas enregistrés. Par conséquent, la prochaine fois que vous ouvrirez et exécuterez le projet, le système vous demandera de vous connecter, sauf si votre administrateur a configuré la connexion unique pour IBM SPSS Modeler Server et les connexions de base de données ODBC. Pour plus d'informations, voir la section *Security Considerations* du guide d'installation d'IBM Analytical Decision Management.

Nom de la table. Sélectionnez la table qui contient les données à utiliser.

Supprimer les espaces de début et de fin. Cochez cette case pour que les espaces situés en début et en fin des chaînes soient supprimés lors de l'importation. Vous pouvez sélectionner la suppression des espaces à gauche, à droite, des deux côtés ou aucun.

Entourer de guillemets les noms des tables et colonnes. Indiquez si vous souhaitez que les noms des tables et des colonnes soient placés entre guillemets lors de l'envoi des requêtes à la base de données (par exemple, s'ils contiennent des espaces ou des signes de ponctuation).

- Si vous sélectionnez **Si nécessaire**, les noms des tables et des champs seront placés entre guillemets *uniquement* s'ils contiennent des caractères non standard. Les caractères non standard incluent les caractères non ASCII, les caractères espace et tout caractère non alphanumérique autre qu'un point (.).
- Sélectionnez **Toujours** si vous souhaitez que *tous* les noms des tables et des champs soient mis entre guillemets.
- Sélectionnez **Jamais** si vous ne souhaitez *jamais* mettre les noms des tables et des champs entre guillemets.

Indiquer les champs d'entrée. Pour plus d'informations, voir la rubrique «Sélection de champs d'entrée», à la page 45.

IBM Analytical Decision Management prend en charge les mêmes sources de données de base de données qu'IBM SPSS Modeler pour la lecture, l'écriture et l'insertion de données. Pour plus d'informations, voir le rapport de compatibilité du logiciel pour IBM SPSS Modeler 18. Ce comportement est différent de celui des bases de données prises en charge par IBM SPSS Collaboration and Deployment Services 8, qui est la base de données hébergeant le référentiel.

Source Vue de données

Une source Vue de données crée une connexion à utiliser avec des règles IBM Operational Decision Management. Les administrateurs peuvent restreindre l'utilisation de règles complexes qui emploient un modèle de données hiérarchique car les règles de ce type peuvent dégrader les performances. Pour plus de détails, voir «Verrouillage des options de données», à la page 85.

Un *noeud de vue de données* peut être configuré dans IBM SPSS Modeler pour inclure dans les flux des données définies dans une vue de données analytique IBM SPSS Collaboration and Deployment Services. Une vue de données analytique définit une structure d'accès aux données qui décrit les entités utilisées dans les modèles prédictifs et les règles métier. La vue associe la structure de données aux sources de données physiques pour l'analyse.

Pour plus d'informations, consultez la documentation d'IBM SPSS Modeler et la section *Source, Process, and Output Nodes* de l'aide d'IBM SPSS Modeler.

Lorsque vous sélectionnez une source Vue de données, vous êtes invité à entrer d'autres détails :

Source de données d'analyse. Sélectionnez la source de données d'analyse dans le référentiel.

Nom de la table. Sélectionnez la table qui contient les données à utiliser. Cette table provient du modèle de données dans la vue de données analytique. Chaque table du modèle de données représente un concept ou une entité impliquée dans le processus d'analyse prédictive. Les champs des tables correspondent aux attributs des entités représentées par les tables. Par exemple, si vous analysez des commandes client, votre modèle de données peut inclure une table pour les clients et une table pour les commandes. La table de clients peut contenir des champs pour l'identificateur, l'âge, le sexe, l'état matrimonial et le pays de la résidence du client. La table de commandes peut contenir des champs pour l'identificateur de la commande, le nombre d'éléments dans celle-ci, le coût total et l'identificateur du client qui a passé la commande. Le champ d'ID client permet d'associer les clients de la table de clients avec les commandes de la table de commandes.

Plan d'accès aux données. Sélectionnez le plan d'accès aux données qui contient les données à utiliser. Un plan d'accès aux données associe les tables de modèle de données d'une vue de données analytique aux sources de données physiques. Une vue de données analytique contient généralement plusieurs plans d'accès aux données. Quand vous changez le plan d'accès aux données utilisé, vous changez les données utilisées par votre projet. Par exemple, si la vue de données analytique contient un plan d'accès aux données d'apprentissage d'un modèle et un plan d'accès aux données de test de modèle, vous pouvez passer des données d'apprentissage aux données de test en modifiant le plan utilisé.

Attributs facultatifs. Si cette option est disponible, vous pouvez sélectionner des attributs facultatifs à ajouter à la source de données.

Indiquer les champs d'entrée. Pour plus d'informations, voir la rubrique «Sélection de champs d'entrée», à la page 45.

Source IBM Cognos BI

Vous pouvez importer des données de base de données IBM Cognos BI ou des rapports de type Liste en vue de leur utilisation dans IBM Analytical Decision Management.

Remarque : Cette option est indisponible si le serveur IBM SPSS Modeler Server utilisé avec IBM Analytical Decision Management ne prend pas en charge le serveur IBM Cognos BI.

Lorsque vous sélectionnez **Serveur IBM Cognos BI** et que vous cliquez sur **Connecter**, vous êtes invité à entrer les détails de connexion du serveur.

1. Entrez l'**URL de serveur** du serveur IBM Cognos à partir duquel vous souhaitez importer ou exporter les données. Contactez votre administrateur système Cognos si vous ne savez pas quel URL choisir.

2. Sélectionnez votre **Mode** de connexion. Pour vous connecter en tant qu'utilisateur spécifique (par exemple, en tant qu'administrateur), sélectionnez **Définir les informations d'identification** et entrez l'espace-noms IBM Cognos, le nom d'utilisateur et le mot de passe :
 - Entrez l'**Espace de noms** du fournisseur d'authentification de sécurité IBM Cognos pour vous connecter au serveur. Le fournisseur d'authentification permet de définir et de gérer les utilisateurs, les groupes et les rôles, ainsi que de contrôler le processus d'authentification.
 - Entrez le **Nom d'utilisateur** sous lequel vous souhaitez vous connecter au serveur.
 - Entrez le **Mot de passe** associé au nom d'utilisateur indiqué.

OU

Parcourez et sélectionnez les données d'identification IBM Cognos stockées dans IBM SPSS Collaboration and Deployment Services Repository.

3. Vous pouvez également vous connecter sans les données d'identification utilisateur. Dans ce cas, vous n'indiquez pas l'espace de noms, le nom d'utilisateur ou le mot de passe, mais vous sélectionnez **Utiliser la connexion anonyme**.

Remarque : certaines connexions serveur n'autorisent pas les connexions anonymes.

Sélection des détails d'objet Cognos

Lorsque vous avez configuré la connexion au serveur IBM Cognos, sélectionnez le mode. Le mode **Données** permet de rechercher et de sélectionner le package à partir duquel les données doivent être importées. Le mode **Rapport de liste** permet de rechercher et de sélectionner le rapport de type Liste à partir duquel les données doivent être importées.

Un package contient un modèle IBM Cognos ainsi que les dossiers, les requêtes, les rapports, les vues, les raccourcis, les URL et les définitions de travail associés à ce modèle. Un modèle IBM Cognos définit des règles métier, des descriptions de données, des relations de données, des dimensions métier et des hiérarchies ainsi que d'autres tâches administratives.

1. Pour sélectionner un package, sélectionnez le mode **Données** et cliquez sur **Parcourir**. Une boîte de dialogue s'affiche, répertoriant tous les packages disponibles à partir desquels vous pouvez importer du contenu.
2. Vous pouvez également sélectionner un rapport de type Liste. Dans ce cas, sélectionnez le mode **Rapport de liste** et cliquez sur **Parcourir**. Une boîte de dialogue s'affiche, répertoriant les packages et les dossiers dans lesquels vous pouvez sélectionner un rapport. *Remarque* : Seuls les rapports de liste unique peuvent être importés, les listes multiples ne sont pas prises en charge.
3. Si nécessaire, développez la liste pour rechercher le package ou le rapport voulu. Sélectionnez l'élément et cliquez sur **OK**. Les détails du package ou du rapport s'affichent dans l'éditeur de source de données.
4. Une fois que vous avez sélectionné un package, sélectionnez les objets à importer. Par exemple, sélectionnez les sujets de requête d'objet package (tables de base de données) ou des éléments de requête de champ (colonnes de table) à utiliser à partir d'un package. *Remarque* : les données doivent être au format UTF-8.
5. Sélectionnez l'option **Regrouper les données avant l'importation** si vous voulez importer des données agrégées plutôt que des données brutes. En général, les données agrégées comportent beaucoup moins d'enregistrements que les données brutes.

Sélectionner les objets. Affiche les objets (également désignés "espaces de noms") sélectionnables dans le package ou le rapport de type Liste. Développez un ou plusieurs objets pour afficher les champs d'entrée que vous pouvez importer. La sélection d'un objet sélectionne automatiquement tous ses champs d'entrée.

Réviser les paramètres. Si des paramètres ont été définis pour un objet données ou pour un rapport, vous pouvez préciser leurs valeurs avant d'importer l'objet ou le rapport. Cette table contient la liste de paramètres obligatoires pour les filtres et les calculs. Ces paramètres peuvent être associés à des valeurs par défaut. Un avertissement apparaît lorsqu'un ou plusieurs paramètres n'ont pas de valeur par défaut.

Définissez les paramètres comme il convient dans la colonne **Valeur**. Vous pouvez employer des paramètres pour procéder à des ajustements (par exemple, pour effectuer un calcul paramétré) avant d'importer les données.

Réviser les filtres. Si un package est associé à des filtres définis, vous pouvez en importer un ou plusieurs. Si le filtre que vous importez est associé à des données importées, il est appliqué avant l'importation des données. Les filtres peuvent être paramétrés.

Vérifier les champs d'entrée. Répertoriez les champs d'entrée (également désignés "objets de base de données") qui seront importés dans IBM Analytical Decision Management pour traitement. Si vous n'avez pas besoin d'un champ particulier, décochez la case située en regard du nom de champ correspondant.

Icônes d'objet Cognos : Les divers types d'objets pouvant être importés depuis une base de données Cognos BI sont représentés par différentes icônes, comme l'illustre le tableau suivant.

Tableau 2. Icônes d'objet Cognos.

Icône	Objet
	Package
	Espace de noms
	Objet de requête
	Élément de requête
	Dimension de mesure
	Mesure
	Dimension
	Hiérarchie de niveau
	Niveau
	Filtre
	Rapport
	Calcul autonome

Source Analytic Server

Vous ne pouvez pas sélectionner le type de source de données du serveur analytique lors de la création d'une nouvelle source de données. Cette information est uniquement disponible pour affichage lorsque vous utilisez IBM SPSS Modeler Advantage pour ouvrir un modèle créé dans IBM SPSS Modeler d'après une source de données Analytic Server.

IBM SPSS Analytic Server a été introduit dans IBM SPSS Modeler 16, ce qui permet d'utiliser HDFS (Hadoop Distributed File System) comme source de données. Les utilisateurs d'IBM SPSS Modeler peuvent créer des flux à l'aide du noeud source HDFS Analytic Server et générer des modèles d'après le système de volumes massifs de données (big data) Hadoop. Les utilisateurs d'IBM Analytical Decision

Management peuvent ouvrir ces flux de modèles dans IBM SPSS Modeler Advantage (ils ne peuvent pas créer de projets à l'aide de sources de données Analytic Server). Vous ne pouvez pas exécuter certaines actions telles que la modification d'une cible ou d'un type de modèle.

Lorsque vous affichez une source de données Analytic Server dans l'onglet Données, les informations suivantes sont disponibles :

Nom de la source de données. Nom attribué à la source de données dans IBM SPSS Modeler. Vous pouvez, si vous le souhaitez, modifier ce nom. Les autres champs sont non modifiables.

Source **Analytic Server.** Chemin d'accès à la source de données tel que configuré dans IBM SPSS Modeler.

Indiquer les champs d'entrée. Les champs d'entrée utilisés.

Sélection de champs d'entrée

Lorsque vous disposez d'un grand nombre de données, vous pouvez vouloir réduire ou affiner les données importées depuis votre source (par exemple, pour n'utiliser que les champs pertinents pour votre centre d'intérêt.

1. En bas de la boîte de dialogue Editeur de source de données, cliquez sur **Indiquer les champs d'entrée** ; la boîte de dialogue est développée pour afficher tous les champs de données de votre source. En outre, toutes les expressions que vous créez sont également affichées. Pour plus d'informations sur la création d'expressions, voir «Générateur de formules», à la page 49.
2. Sélectionnez chacun des champs et expressions à utiliser.

Outre la sélection de données, vous pouvez, si nécessaire, changer le niveau de mesure d'un ou de plusieurs champs de données. Pour plus d'informations, voir «Niveaux de mesure», à la page 46.

Aperçu des données

Pour vérifier que vous avez sélectionné la source de données appropriée, cliquez l'icône d'**Aperçu** pour afficher un échantillon des données que contient la source.

Présentation des données

La fonction de présentation des données est disponible sur l'onglet Données de toutes les applications, sur les onglets Modélisation et Evaluation d'IBM SPSS Modeler Advantage, ainsi que sur l'onglet Déployer des applications qui incluent la fonction *Evaluer maintenant*.

1. Pour exécuter une présentation de source de données, cliquez sur l'icône Présentation des données disponible dans toute l'application.



Figure 11. Icône Présentation des données

2. La boîte de dialogue Présentation des données s'affiche. Si vous le souhaitez, sélectionnez un champ de superposition dans la liste déroulante de la section Options de présentation.
Ensuite, une fois la présentation exécutée, des onglets sont disponibles pour afficher uniquement les résultats du premier champ sélectionné ou pour superposer celui-ci avec le champ spécifié dans ce menu déroulant. Par exemple, vous pouvez souhaiter extraire des résultats des informations sur l'âge des clients, puis superposer un autre champ tel que le **genre**.
3. Si vous le souhaitez, sélectionnez **Utiliser les données partitionnées** dans la section Options de présentation. Cette option est disponible si, dans l'onglet Modélisation, sous Paramètres facultatifs, vous avez sélectionné l'option **Partitionner automatiquement les données pour activer le scoring du modèle dans la source de données de génération pour le scoring et le test**. La sélection de cette

option divise les données en sous-ensembles ou échantillons distincts à des fins d'apprentissage et de test du modèle. En générant le modèle sur un sous-ensemble et en le testant sur un autre, vous pouvez avoir une idée de la façon dont il va se généraliser à d'autres ensembles de données.

4. Sélectionnez les champs de données à inclure dans la présentation et cliquez sur **Exécuter la présentation**. Tous les champs disponibles dans la source de données sont répertoriés. Par défaut, tous les champs d'entrée du modèle et la cible sont sélectionnés.

La présentation de données s'exécute et les résultats apparaissent. Vous pouvez trier les informations et sélectionner les colonnes à afficher.

5. Cliquez sur un champ pour en visualiser les détails. Un nouvel onglet Résultats s'ouvre pour chaque champ sélectionné, ce qui vous permet de visualiser des graphiques et des tables, ainsi que sélectionner des champs de superposition si cette fonction est disponible.

Niveaux de mesure

Le niveau de mesure aide à caractériser les informations qui sont représentées par chaque champ de données. Il peut déterminer le mode d'utilisation d'un champ particulier dans les règles, dans la modélisation ou dans d'autres applications. Le niveau de mesure peut être indiqué dans le modèle de données du projet pour une source de données ; par exemple, vous pouvez définir sur *Indicateur* le niveau de mesure d'un champ d'entier comportant les valeurs 0 et 1 pour indiquer que 0 = *False* et 1 = *True*. Vous pouvez également changer le niveau de mesure dans la boîte de dialogue **Editeur de source de données** lorsque vous indiquez les champs d'entrée à utiliser. Pour plus d'informations, voir «Sélection de champs d'entrée», à la page 45.

Les niveaux de mesure suivants sont disponibles :

- **Continue** Utilisé pour décrire des valeurs numériques, telles qu'une plage de 0 à 100 ou de 0.75 à 1.25. Une valeur continue peut être un entier, un nombre réel ou une date/heure.
- **Catégorielle** Utilisé pour des valeurs de chaînes lorsque le nombre exact de valeurs distinctes est inconnu. Il s'agit d'un type de données **non instancié**, ce qui signifie que toutes les informations possibles sur le stockage et l'utilisation des données ne sont pas encore connues. Une fois les données lues, le niveau de mesure sera *Indicateur*, *Nominal* ou *Sans type*, en fonction du nombre maximal spécifié de membres pour les champs nominaux dans la boîte de dialogue Propriétés du flux.
- **Indicateur** Utilisé pour des données ayant deux valeurs distinctes qui indiquent la présence ou l'absence d'une caractéristique, telle que vrai et faux, Oui et Non ou 0 et 1. Dans IBM Analytical Decision Management, la première valeur de la liste de modalités sera utilisée comme valeur « faux » et la deuxième sera la valeur « vrai », comme indiqué par une infobulle lorsque vous passez au-dessus des valeurs qui s'affichent dans l'onglet Données. Vous pouvez représenter les données sous forme de texte, d'entier, de nombre réel, de date, d'heure ou d'horodatage.
- **Nominal** Utilisé pour décrire des données ayant plusieurs valeurs distinctes, chacune étant traitée en tant que membre d'un ensemble (par exemple *small/medium/large*). Les données nominales peuvent bénéficier de n'importe quel stockage numérique, chaînes ou date/heure. Le fait de définir le niveau de mesure sur *Nominal* n'a pas pour effet de convertir automatiquement les valeurs en stockage de chaîne. Pour plus d'informations sur la définition du nombre maximal de membres autorisés pour les champs nominaux, voir «Propriétés», à la page 33.
- **Ordinal** Utilisé pour décrire des données avec plusieurs valeurs distinctes ayant un ordre inhérent. Par exemple, les catégories de salaire ou l'indice de satisfaction peuvent avoir le type données ordinales. L'ordre est défini par l'ordre de tri naturel des éléments des données. Par exemple, 1, 3, 5 est l'ordre de tri par défaut d'un ensemble d'entiers, alors que *ELEVE*, *FAIBLE*, *NORMAL* (tri alphabétique croissant) est l'ordre d'un ensemble de chaînes. Le niveau de mesure ordinal vous permet de définir un ensemble de données catégorielles comme des données ordinales, pour la visualisation, la création de modèles et l'exportation vers d'autres applications (telles qu'IBM SPSS Statistics) qui reconnaissent les données ordinales comme un type distinct. Vous pouvez utiliser le champ ordinal partout où un champ nominal peut être utilisé. De plus, les champs de n'importe quel type de stockage (réel, entier, chaîne, date, heure, etc.) peuvent être définis comme ordinaux.

- **Sans type** Utilisé pour des données qui ne sont pas conformes aux types ci-dessus, pour des champs à valeur unique ou pour des données nominales où l'ensemble contient plus de membres que le maximum défini. Ce type s'avère pratique dans les cas où autrement le niveau de mesure serait un ensemble avec de nombreux membres (par exemple, un numéro de compte). Lorsque vous sélectionnez **Sans type** pour un champ, le rôle est automatiquement défini sur **Aucun**, avec l'**ID d'enregistrement** comme seule alternative. La taille maximale par défaut des ensembles est de 250 valeurs uniques. Ce nombre peut être ajusté ou désactivé dans la boîte de dialogue Propriétés du projet, accessible via l'icône de la barre d'outils.
- **Collection** Utilisé pour identifier des données non géospatiales enregistrées dans une liste. Une collection est effectivement un champ de liste de profondeur nulle, dont les éléments ont l'un des autres niveaux de mesure.
- **Géospatial** Utilisé avec le type de stockage Liste pour identifier des données géospatiales. Ces listes peuvent être des champs Liste de nombres entiers ou Liste de nombres réels dont la profondeur est comprise entre zéro et deux, inclus.

Edition des valeurs de champ de données

Dans les tables de données, la colonne de **Valeurs** répertorie les valeurs de chaque champ de la source de données (par exemple, valeurs maximales et minimales d'une plage). Pour les champs d'indicateur, placez le pointeur de la souris sur la valeur pour afficher une infobulle qui indique les valeurs vraies ("true") et fausses ("false") définies pour le champ. Vous pouvez également cliquer sur une valeur pour l'éditer. Une boîte de dialogue dont le contenu varie selon le type de champ de données sélectionné s'affiche. Par exemple, si le type de champ est :

- **Continu**, vous pouvez entrer des valeurs limites inférieure et supérieure ;
- **Indicateur**, vous pouvez permuter les valeurs True (vrai) et False (faux) ;
- **Nominal**, vous pouvez ajouter, supprimer, réorganiser et renommer des valeurs.

Mappage de champs

Lorsque vous ajoutez une nouvelle source de données, vous pouvez devoir mapper tous ses champs pour garantir sa compatibilité avec le modèle de données du projet. Par exemple, si le modèle de données du projet nécessite un champ appelé *achat* avec des valeurs *Oui* et *Non* (indicateur de niveau de mesure), toute source de données utilisée doit comporter un champ compatible. Si les noms de champ ne sont pas identiques, ils peuvent être mappés en conséquence. Le champ d'entrée et le champ mappé associé doivent présenter le même type de données.

Si le modèle de données du projet change, le mappage entre celui-ci et de nombreuses sources de données peut devenir incorrect. Si tel est le cas, dans la section **Sources de données du projet**, l'indicateur de compatibilité (boule) change de couleur en conséquence pour chacune de ces sources de données ; vous devez alors les remapper avec le nouveau modèle de données du projet.

1. Dans la section **Sources de données du projet**, cliquez sur **Mapper les champs**. Une boîte de dialogue s'affiche. Le système effectue une meilleure supposition en mappant les champs pour vous. A tout moment, vous pouvez revenir à la meilleure supposition du système en cliquant sur **Réinitialiser**.
2. Sélectionnez un champ dans la colonne **Entrées disponibles** et le champ correspondant dans la colonne **Entrées requises**, puis cliquez sur **Mapper** pour mapper les deux champs. Répétez l'opération jusqu'à ce que tous les champs de la colonne **Entrées disponibles** soient mappés.
3. Lorsque vous avez terminé, cliquez sur **OK**.

Lorsque c'est possible, IBM Analytical Decision Management mappe automatiquement les types de stockage compatibles quand un utilisateur ajoute une source de données supplémentaire. Si le type n'est pas compatible, l'utilisateur sera autorisé à convertir les types.

La table suivante répertorie les mappages autorisés.

Tableau 3. Types qui peuvent être convertis

A →						
De ↓	Entier	Réel	Chaîne	Date	Heure	horodatage
Entier	n/a	Auto	Oui	Non autorisé	Non autorisé	Non autorisé
réel	Oui (décimaux tronqués)	n/a	Oui	Non autorisé	Non autorisé	Non autorisé
Chaîne	Oui	Oui	n/a	Oui	Oui	Oui
Date	Non autorisé	Non autorisé	Oui	n/a	Non autorisé	Oui (heure définie sur 00:00)
Heure	Non autorisé	Non autorisé	Oui	Non autorisé	n/a	Oui (date de base)
Horodatage	Non autorisé	Non autorisé	Oui	Auto (heure tronquée)	Auto (date tronquée)	n/a

Calcul des champs

Dans la plupart des applications IBM Analytical Decision Management, vous pouvez ajouter des champs au modèle de données du projet, tels que des formules, des sorties de modèle, des champs créés à partir de règles de segment ou des champs de tables secondaires. L'onglet **Dérivé** de l'onglet Données affiche tous les champs à inclure en tant qu'élément des données, et les champs peuvent être ordonnés de façon à pouvoir être utilisés cumulativement (leur ordre dans la table correspondra à l'ordre dans lequel ils sont calculés ; par exemple, le deuxième champ dérivé peut utiliser le champ dérivé en premier comme entrée).

L'onglet **Dérivé** n'est pas inclus dans l'application IBM SPSS Rules Management.

- Dans l'onglet Données, cliquez sur l'onglet **Dérivé**.

Pour ajouter une expression :

- Cliquez sur l'icône **Dériver le champ à l'aide d'une expression**. Le Générateur de formules s'affiche. Pour plus d'informations, voir «Générateur de formules», à la page 49.

Pour ajouter une sortie de modèle :

1. Cliquez sur l'icône **Dériver le champ à l'aide d'un modèle ou d'une règle**, puis recherchez et sélectionnez un modèle. La boîte de dialogue **Entrées et sorties** s'affiche pour le modèle sélectionné.
2. Si les champs d'entrée du modèle sont compatibles, la section **Mappage d'entrée du modèle** est réduite dans la boîte de dialogue. Si les champs d'entrée du modèle ne sont pas compatibles, la section est développée. Vous devez mapper les entrées de table disponibles avec les entrées du modèle sélectionné.
3. Dans la section de **Sorties du modèle**, toutes les sorties du modèle disponibles pour le modèle sélectionné sont répertoriées. Sélectionnez les champs de sortie du modèle à ajouter à vos données. La colonne **Nom de sortie du modèle** affiche le nom du champ de sortie généré par le modèle.

Tout champ produit par le modèle peut être sélectionné. Les champs de sortie du modèle sont les champs sortants du flux qui ne servent pas également d'entrée pour ce flux (champs ajoutés par le flux). Les champs d'entrée peuvent ne pas inclure tous les éléments du modèle de données du projet. Par conséquent, un champ de sortie de modèle peut porter le même nom qu'un champ de modèle de données du projet. Ainsi, vous pouvez également changer les noms de champ de données.

La colonne **Mesure** affiche le type de mesure de la sortie de modèle, préfixé avec l'icône de type approprié. La colonne **Valeurs** affiche les valeurs de la sortie de modèle. Ces colonnes peuvent afficher les résultats de certains champs uniquement. En général, les types et les valeurs de mesure

sont connus pour les sorties de modèle réelles (telles que les valeurs prévues, les probabilités) sans analyse des données et, par conséquent, sont affichés. Cependant, les types ou les valeurs de mesure peuvent être inconnus pour d'autres champs de sortie d'un flux (par exemple, un champ dérivé) sans analyse de données.

4. Une section **Paramètre du modèle** sera incluse dans la partie inférieure de la boîte de dialogue, si le flux sélectionné comprend des paramètres d'évaluation. Entrez les valeurs des paramètres du modèle.
5. Cliquez sur **Enregistré** dès que vous avez terminé avec la boîte de dialogue. Toutes les sorties de modèle sélectionnées sont ajoutées à la section **Champs supplémentaires** de l'onglet Dérivé.

Pour créer un champ à l'aide de règles de segment :

1. Cliquez sur l'icône **Dériver le champ à l'aide des règles de segment**.
2. Créez une règle de segment. Les segments sont définis par une ou plusieurs expressions qui retournent un résultat true (vrai) ou false (faux). Pour plus d'informations, voir «Définition de règles de segment», à la page 53.
3. Cliquez sur **OK** lorsque vous avez terminé. La nouvelle règle de segment est ajoutée à la section de **Champs supplémentaires** de l'onglet Dérivé.

Pour ajouter un champ depuis une table secondaire :

1. Si vous avez défini une table secondaire avec une relation linéaire sur l'onglet Données, cliquez sur l'icône **Ajouter le champ depuis une table secondaire**.
2. Sélectionnez une table secondaire dans le menu déroulant, puis sélectionnez un ou plusieurs champs de table.
3. Cliquez sur **Enregistré** dès que vous avez terminé avec la boîte de dialogue. Le(s) champ(s) est/sont ajouté(s) à la section de **Champs supplémentaires** de l'onglet Dérivé.

Générateur de formules

Vous pouvez utiliser le générateur de formules pour dériver des champs ou des attributs supplémentaires à l'usage de l'application. Par exemple, si vous utilisez des données bancaires, vous pouvez créer une expression qui affiche le rapport entre le revenu d'un client et le nombre de comptes de prêt détenus.

Les expressions sont disponibles dans toute l'application de la même manière que les données opérationnelles (par exemple, dans les règles et les prédicteurs de modèle).

Pour afficher des expressions existantes, ou créer une nouvelle expression, accédez à l'onglet Données, puis cliquez sur l'onglet **Dérivé**.

Création d'une expression

1. Dans la boîte de dialogue Générateur de formules, entrez un nom unique pour l'expression.
2. Renseignez le champ **Equation d'expression** en vous servant des listes de fonctions et de champs comme références.
ou
3. Sélectionnez les champs et fonctions souhaités dans les listes.
4. Effectuez un double-clic ou cliquez sur la flèche bleue pour ajouter le champ ou la fonction dans le champ **Equation d'expression**.
5. Utilisez les boutons d'opérande situés au centre de la boîte de dialogue pour insérer des opérations dans l'expression.
6. Entrez une **Description** qui explique la fonction de l'expression.
7. Cliquez sur **Vérifier** pour valider l'expression, puis cliquez sur **Enregistrer**.

Remarque : Vous pouvez copier-coller les expressions du client IBM SPSS Modeler, mais notez que les résultats peuvent ne pas toujours être ceux attendus si l'expression est dépendante du flux (par exemple,

si vous collez une fonction @ ou de ligne prise en charge, elle peut ne pas produire les résultats souhaités car le reste du flux peut ne pas être configuré pour la prendre en charge).

Seules les expressions fournies dans l'interface utilisateur d'IBM Analytical Decision Management sont officiellement prises en charge.

Sélection de fonctions

La liste des fonctions affiche toutes les fonctions et tous les opérateurs disponibles. Faites défiler pour sélectionner une fonction dans la liste ou, pour faciliter une recherche, utilisez la liste déroulante pour afficher un sous-ensemble de fonctions et d'opérateurs. Les fonctions disponibles sont regroupées en catégories pour faciliter vos recherches.

Voici des détails concernant les catégories. Pour plus d'informations, voir la documentation IBM SPSS Modeler.

Tableau 4. Catégories de fonctions

Type de fonction	Description
Fonctions utilisées récemment	Contient la liste des fonctions utilisées dans la session en cours.
Opérateurs	Répertorie tous les opérateurs disponibles pour la génération d'expressions. Les opérateurs sont également disponibles via les boutons situés au centre de la boîte de dialogue.
Fonctions d'informations	Servent à obtenir un aperçu des valeurs de champ. Par exemple, la fonction <code>is_string</code> renvoie true pour tous les enregistrements de type chaîne (string).
Fonctions de conversion	Servent à créer des champs ou à convertir un type de stockage. Par exemple, la fonction <code>to_timestamp</code> convertit le champ sélectionné en un horodatage.
Fonctions de comparaison	Servent à comparer des valeurs de champ, entre elles ou par rapport à la chaîne spécifiée. Par exemple, <code><=</code> est utilisé pour vérifier si les valeurs de deux champs sont inférieures ou égales.
Fonctions logiques	Servent à exécuter des opérations logiques, comme <code>if</code> , <code>then</code> et <code>else</code> .
Fonctions numériques	Servent à effectuer des calculs numériques, comme le calcul du logarithme naturel de valeurs de champ.
Fonctions aléatoires	Sert à sélectionner des éléments de manière aléatoire ou pour générer des nombres.
Fonctions sur chaînes	Sert à effectuer diverses opérations sur des chaînes, telles que <code>stripchar</code> , qui permet de supprimer un caractère spécifié.
Fonctions date/heure	Sert à effectuer diverses opérations sur des champs de date, d'heure et d'horodatage.
Fonctions nul et non renseigné	Servent à atteindre, à marquer et souvent à remplir des blancs définis par l'utilisateur ou des valeurs système manquantes. <code>@BLANK</code> n'est pas pris en charge (car son comportement dépend d'autres noeuds, comme le noeud Type). <code>@NULL(FIELD)</code> et <code>undef</code> sont pris en charge. <code>@NULL(FIELD)</code> est de type booléen et retourne la valeur vraie (True) si la valeur de FIELD est la valeur système manquante <code>\$null\$</code> . Renvoie une valeur fausse (False) pour toutes les autres valeurs, notamment les blancs définis par l'utilisateur. Si vous souhaitez vérifier les deux, utilisez <code>@BLANK(FIELD)</code> et <code>@NULL(FIELD)</code> . <code>undef</code> est de type Any et sert généralement à saisir une valeur <code>\$null\$</code> . Par exemple, pour remplir des blancs avec des valeurs nulles dans le noeud Filler.
Toutes les fonctions	Contient la liste complète des fonctions disponibles.

Sélection de champs et de valeurs

La liste des champs affiche tous les champs disponibles. Sélectionnez un champ dans la liste. Effectuez un double-clic ou cliquez sur la flèche bleue pour ajouter un champ à l'expression.

Affichage et sélection de valeurs

Les valeurs des champs sont visualisables à partir de nombreux emplacements de l'application. Pour afficher les valeurs d'un champ à partir du générateur de formules, sélectionnez le champ souhaité, puis cliquez sur le bouton d'entrées d'expression pour ouvrir une boîte de dialogue qui répertorie les valeurs du champ sélectionné. Vous pouvez ensuite sélectionner une valeur et cliquer sur **Insérer** pour coller la valeur dans l'expression en cours.



Figure 12. Bouton d'entrées d'expression

Pour les champs indicateurs et les champs nominaux, toutes les valeurs définies sont présentées. Pour les champs d'intervalles continus (intervalle numérique), les valeurs minimales et maximales sont affichées.

Vérification des expressions

Cliquez sur **Vérifier** dans le Générateur de formules (angle inférieur droit) pour valider l'expression. Expressions qui n'ont pas encore été vérifiées et affichées en rouge. En cas d'erreurs, un message indiquant la cause s'affiche.

La vérification porte sur les éléments suivants :

- Utilisation correcte des guillemets dans les noms de valeurs et de champs
- Utilisation correcte des paramètres et des variables globales
- Utilisation correcte des opérateurs
- Existence des champs référencés
- Existence et définition des valeurs globales référencées

Si votre syntaxe comporte des erreurs, essayez de recréer l'expression en utilisant les listes et les boutons des opérateurs au lieu de taper l'expression manuellement. Cette méthode ajoute automatiquement les guillemets corrects pour les champs et les valeurs.

Veillez noter les limitations suivantes lorsque vous générez des expressions dans IBM Analytical Decision Management. Les expressions ne peuvent contenir aucun des articles suivants :

- Une référence à un paramètre de flux IBM SPSS Modeler
- Une référence à un flux IBM SPSS Modeler global
- Une référence à une fonction de base de données
- Une référence à une fonction @ de champ ou de valeur de champ spécial suivante :
 - @TARGET
 - @PREDICTED
 - @FIELD
 - @PARTITION_FIELD
 - @TRAINING_PARTITION
 - @TESTING_PARTITION
 - @VALIDATION_PARTITION

Utilisation des expressions qui renvoient un booléen

Le type de données booléen n'existe pas dans IBM SPSS Modeler et IBM Analytical Decision Management. Les expressions booléennes ne peuvent pas être utilisées pour produire directement une nouvelle zone. L'utilisation d'expressions booléennes se limite aux syntaxes d'expression qui attendent une expression booléenne (comme l'instruction `if`). Voici trois exemples :

```
if ('RFM Score' > 5) then "T" else "F" endif
if ('RFM Score' < 5) then "Low" elseif ('RFM Score' >= 5 and 'RFM Score' < 10) then "Medium" else "High" endif
hassubstring('Phone type','Coolphone')
```

Dans les deux premiers exemples ci-dessus, les expressions relatives à "l'évaluation RFM" ('RFM Score' > 5, 'RFM Score' >= 5, et 'RFM Score' < 10) peuvent être considérées comme booléennes.

Ces expressions ne peuvent pas être utilisées seules pour dériver une zone. Cela aurait pour effet d'afficher une erreur telle que la suivante : Invalid type for an expressions which is not a condition: Boolean (type non valide pour une expression qui n'est pas une condition : booléen). Toutefois, des expressions telles que celles-ci peuvent être utilisées dans des expressions conditionnelles, comme dans les exemples complets, qui elles peuvent dériver des zones à mesures catégoriques, comme la zone de type indicateur. L'indicateur est la mesure par défaut pour une zone à deux valeurs.

Le second exemple ci-dessus permet de générer une zone à trois valeurs distinctes (si tant est que l'acquisition de données trouve des enregistrements appartenant aux catégories Low, Medium et High). Le niveau de mesure de la zone dérivée aurait alors probablement une valeur Nominale par défaut. Le troisième exemple ci-dessus a peu de chances de renvoyer une zone à deux valeurs distinctes.

Pour plus d'informations sur les niveaux de mesure, consultez la section «Niveaux de mesure», à la page 46.

Chapitre 5. Utilisation des règles

Les règles sont utilisées pour automatiser les décisions en fonction d'une logique applicative ou des données de sortie de modèles prédictifs. Selon l'application, vous pouvez utiliser des règles pour exclure des clients en fonction de l'historique des paiements, transférer les demandes d'indemnisation à risque à une unité d'investigation spécialisée ou cibler les personnes associées à la plus forte propension à la résiliation, à l'acquisition ou à la fraude, identifiée par un modèle prédictif. Tandis que les règles apportent l'automatisation, les modèles offrent un pouvoir prédictif. La possibilité de combiner ces deux atouts constitue l'une des fonctions principales qui distinguent les applications prédictives des autres outils d'analyse.

Les règles peuvent être créées, appliquées et enregistrées à différents points dans chaque application. Vous pouvez également créer et éditer des règles partagées avec IBM SPSS Rules Management.

Création des règles

- Dans l'application dans laquelle vous voulez ajouter la règle, cliquez sur l'icône **Créer une règle**.
- Vous pouvez également cliquer sur **Rechercher une règle existante** pour parcourir les règles existantes qui peuvent être ajoutées ou réutilisées dans l'application en cours. Pour plus d'informations, voir la rubrique «Exportation et réutilisation de règles», à la page 56.
- Pour créer une règle partagée, lancez IBM SPSS Rules Management à partir de la page de lancement d'application.

Définition de règles de segment

Des règles de segment sont utilisées dans toute l'application pour sélectionner et affecter des enregistrements dans différentes applications selon les besoins. Les segments sont définis par une ou plusieurs expressions qui retournent un résultat true (vrai) ou false (faux). Par exemple, une règle avec un segment unique peut renvoyer la valeur true (vrai) si la valeur de la zone d'âge est inférieure à 18. Les segments peuvent être déterminés selon les zones ou les expressions définies dans les données ou selon les scores de propension ou les prévisions renvoyées par un modèle. Dans les applications dotées d'un onglet Définir et configurées pour comporter plusieurs hiérarchies de dimension (désignées *éléments*), vous pouvez également sélectionner d'autres membres de dimension.

Chaque segment est défini par une ou plusieurs expressions et renvoie la valeur true si toutes les conditions sont remplies. Par exemple, un segment défini par âge et par genre inclut seulement les enregistrements qui remplissent les deux conditions ; la règle renvoie *false* pour tous les autres enregistrements.

Pour créer ou éditer une règle de segment

1. Dans l'application dans laquelle vous voulez ajouter la règle, cliquez sur l'icône **Créer une règle**.
2. Indiquez un nom pour la règle.
3. Pour la colonne d'entrée, cliquez sur l'icône flèche vers le bas pour indiquer une zone, une règle ou un modèle, selon les besoins.

Tableau 5. Options de colonne d'entrée.

Icône	Utilisation
	Activée, cette fonction permet de sélectionner un membre de dimension à partir d'un élément précédemment défini.
	Permet de sélectionner une zone à partir de l'ensemble de données en cours.

Tableau 5. Options de colonne d'entrée (suite).

Icône	Utilisation
	Permet de sélectionner une règle ou un modèle existant dans le référentiel.
	Permet de générer un nouveau modèle.
	Permet de générer une expression.

4. Sélectionnez la fonction que vous voulez utiliser. Pour obtenir les options disponibles, reportez-vous au tableau suivant.

Tableau 6. Opérateurs disponibles.

Type de valeur	Opérateurs disponibles
Horodatage, Date, Date heure	<ul style="list-style-type: none"> • = (égal à) • != (différent de) • > (supérieur à) • < (inférieur à) • >= (supérieur ou égal à) • <= (inférieur ou égal à) • ENTRE • EST NULL • N'EST PAS NULL
Plage	<ul style="list-style-type: none"> • = (égal à) • != (différent de) • > (supérieur à) • < (inférieur à) • >= (supérieur ou égal à) • <= (inférieur ou égal à) • ENTRE • EST NULL • N'EST PAS NULL • PARMI • SAUF
Ensemble	<ul style="list-style-type: none"> • = (égal à) • != (différent de) • EST NULL • N'EST PAS NULL • PARMI • SAUF
Indicateur	<ul style="list-style-type: none"> • = (égal à) • != (différent de) • EST NULL • N'EST PAS NULL

5. Indiquez une valeur appropriée pour le type d'entrée. Pour les valeurs non numériques, une liste déroulante d'options est fournie.

Vous pouvez également cliquer sur **Rechercher une règle existante** pour explorer les règles déjà définies. Pour plus d'informations, voir la rubrique «Exportation et réutilisation de règles», à la page 56.

Une règle de segment comme Age ENTRE 50 et 20 ne renverra pas des résultats. Elle équivaut à l'expression $x \leq \text{valeur}$ et $\text{valeur} < y$, qui ne peut jamais être vraie. Cette règle est cohérente par rapport au comportement de SQL Server. Age ENTRE 20 et 50 est le format correct.

Veillez également noter qu'une règle de segment comme Gender = \$null\$ ne renverra pas de résultats non plus. Si vous souhaitez exclure les champs pour lesquels la valeur Gender est vide, vous pouvez à la place utiliser une expression comme Gender N'est pas parmi (Male, Female).

Jeux de règles de segment

- Les règles de segment peuvent également être combinées en jeux. Un jeu de règles de segment renvoie la valeur true si *l'un* des différents segments est vrai, contrairement aux expressions *d'un* segment, qui doivent toutes être vraies.
- Vous pouvez exporter des jeux de règles de segment à partir de toute application qui utilise des règles. Vous pouvez également le faire dans IBM SPSS Rules Management si cette fonction est disponible. Pour plus d'informations, voir la rubrique «Exportation et réutilisation de règles», à la page 56.

Définition de règles de sélection

Une règle de sélection est un ensemble d'une ou de plusieurs règles de segment (ou de jeu de règles) possédant chacune un attribut *Inclure* ou *Exclure*. Les règles de sélection sont utilisées dans les applications pour sélectionner des sous-ensembles d'enregistrements pour la modélisation, l'évaluation et d'autres traitements.

1. Dans l'application dans laquelle vous voulez ajouter la règle, cliquez sur l'icône **Créer une règle**. Cette icône est disponible aux points où des règles de sélection peuvent être appliquées.
2. Indiquez un nom pour la règle, précisez si vous souhaitez utiliser la règle pour l'inclusion ou l'exclusion (ou faites-le ultérieurement) et ajoutez une ou plusieurs expressions selon vos besoins. Pour plus d'informations, voir la rubrique «Définition de règles de segment», à la page 53.
3. Cliquez sur **OK** pour enregistrer la règle.
4. Cliquez sur les icônes dans la colonne **Inclure/Exclure** si vous souhaitez changer ce paramètre à bascule pour chaque règle. Pour les jeux de règles de sélection référencés, ce paramètre ne peut pas être changé. Vous ne pouvez pas combiner des inclusions et des exclusions dans la même règle.

Vous pouvez également cliquer sur **Rechercher une règle existante** pour explorer les règles déjà définies. Pour plus d'informations, voir la rubrique «Exportation et réutilisation de règles», à la page 56.

Evaluation des règles de sélection

Cliquez sur l'icône **Afficher la logique de la règle de sélection** pour voir les détails sur la façon dont les sélections sont effectuées (ET logique / OU).

- Lors de l'évaluation des règles de sélection, si plusieurs segments sont spécifiés, les enregistrements correspondant à *l'un* des segments sont inclus. Par exemple, les clients peuvent être inclus d'après la notation de crédit OU le mandat.
- Si plusieurs segments d'exclusion sont indiqués, les enregistrements correspondant à l'un des segments sont exclus. Par exemple, les clients peuvent être exclus d'après l'âge OU la notation de crédit.
- En cas de conflit, les règles d'exclusion ont la priorité. Par exemple, un client qui correspond à un segment exclu est exclu indépendamment des segments auxquels il est admissible.
- Lors de la création d'un jeu de règles de sélection dans IBM SPSS Rules Management, toutes les règles doivent être du même type : *Include* ou *Exclude*. Vous ne pouvez pas combiner les règles d'exclusion et d'inclusion dans le même jeu de règles.

Exportation et réutilisation de règles

Par défaut, les règles sont enregistrées dans l'application dans laquelle elles sont créées. Elles peuvent également être exportées ou enregistrées dans IBM SPSS Collaboration and Deployment Services Repository en vue d'être utilisées dans plusieurs applications. Dans ce cas, chaque application enregistre une référence à une version spécifique marquée de la règle partagée ; tout changement apporté à la règle est automatiquement appliqué aux applications qui utilisent cette version.

Exportation de règles

A partir de toute application contenant une règle locale à partager, cliquez sur l'icône **Exporter** pour enregistrer la règle dans IBM SPSS Collaboration and Deployment Services Repository, où elle peut être réutilisée si nécessaire.

- Lorsque vous exportez plusieurs règles, indiquez si vous voulez toutes les exporter en un seul jeu de règles ou comme règles individuelles. Dans le cas de règles individuelles, chaque segment est exporté en tant que règle distincte.
- Lorsque vous exportez plusieurs règles de sélection, si vous sélectionnez **Stocker avec - Sélections**, un jeu de règles de sélection est exporté. Dans ce cas, vous pouvez exporter des inclusions ou des exclusions, mais pas les deux. Si vous avez besoin d'une combinaison d'inclusions et d'exclusions, celles-ci doivent être exportées dans des jeux de règles distincts. Si **Stocker avec - Sélections** n'est pas sélectionné, un jeu de règles de segment est exporté.
- Lorsque vous exportez à partir d'un éditeur de règle d'agrégation, si vous sélectionnez **Export with risk points (Exporter avec les points de risque)**, un jeu de règles d'agrégation est créé. Si cette option n'est pas sélectionnée, un jeu de règles de segment est exporté.

Utilisation de règles existantes

- Dans toute application dans laquelle vous souhaitez utiliser une règle partagée, cliquez sur l'icône **Rechercher une règle existante** pour rechercher des règles existantes dans IBM SPSS Collaboration and Deployment Services Repository, réutilisables dans l'application en cours.
- Pour faire référence à une version spécifique marquée, cliquez sur la flèche en regard du nom de la règle et sélectionnez la version souhaitée. Si aucune version n'est sélectionnée, la plus récente est utilisée. (*LATEST* est un libellé spécial automatiquement déployé à chaque nouvelle version créée.) L'utilisation d'une version spécifique marquée est recommandée pour éviter des résultats inattendus lors de la création d'une nouvelle version d'une règle.

Dans de nombreux cas, les règles référencées peuvent être combinées à des règles locales, mais elles se distinguent visuellement de celles-ci par du texte en gras et par une icône qui indique le type de la règle (ou du jeu de règles).

- Les règles référencées peuvent également être converties en règles locales et enregistrées dans l'application en cours. Pour ce faire, cliquez sur le nom d'une règle référencée, puis sélectionnez **Convertir en règles uniques** dans la boîte de dialogue **Aperçu de la règle**.

Instructions pour partager des règles

Les instructions suivantes s'appliquent lorsque vous recherchez ou utilisez des règles partagées :

- Les règles peuvent être exportées à partir de toute application en vue de leur réutilisation, mais une fois exportées, elles ne peuvent être éditées que dans IBM SPSS Rules Management. IBM SPSS Rules Management peut ne pas être installé sur tous les sites ou disponible pour tous les utilisateurs.
- Il est fortement recommandé d'utiliser une version avec libellés spécifique (autre que *LATEST*), en particulier dans un environnement de production, car les changements apportés à la version la plus récente peuvent avoir des conséquences imprévues. Dans certains cas, par exemple lorsque vous faites référence à une règle à partir d'une autre règle, une version marquée est obligatoire.

- Des règles uniques peuvent être ajoutées à partir d'IBM SPSS Collaboration and Deployment Services Repository par référence, ou converties en règles locales à partir de la boîte de dialogue Aperçu de la règle.
- Les règles uniques peuvent être référencées et combinées à des règles locales dans tout éditeur de règle. Exportées en jeu de règles, les références imbriquées à des règles uniques sont conservées.
- Les jeux de règles de segment et de sélection peuvent être référencés et combinés comme des règles uniques dans la plupart des cas. Cependant, le jeu de règles combiné ne peut pas être exporté sans convertir au préalable les jeux de règles imbriqués en règles uniques. Par conséquent, un jeu de règles exporté peut contenir des références à des règles imbriquées, *mais pas à des jeux de règles imbriqués*.
- Les jeux de règles d'agrégation peuvent également être référencés, mais ils ne peuvent pas être combinés à d'autres règles ni contenir de référence imbriquée à d'autres règles. Dans la référence à un jeu de règles d'agrégation, la règle référencée remplace tout le contenu de l'éditeur de règle d'agrégation. (Sinon, en l'absence de référence à un jeu de règles d'agrégation, l'éditeur d'agrégation peut contenir une combinaison de règles locales, de règles de segment référencées et de jeux de règles de segment.)
- Avec IBM SPSS Rules Management, les mêmes restrictions que pour les règles exportées s'appliquent. Les règles imbriquées sont autorisées, mais pas les jeux de règles imbriqués. En effet, IBM SPSS Rules Management est conçu pour créer et éditer des règles partagées (comme pour exporter à partir d'une autre application).
- Pour éviter les restrictions sur les jeux de règles référencés, dans IBM SPSS Rules Management ou ailleurs, un jeu de règles peut toujours être converti en règles locales distinctes.
- Les règles partagées ne peuvent pas contenir de références aux règles IBM Operational Decision Management externes.

Ajout d'annotations

Dans une agrégation ou dans une affectation, cliquez sur l'icône **Ajouter des annotations** pour ajouter une colonne de champs de texte libres qui permettent d'entrer une note à inclure dans les données de sortie de règle. Par exemple, il peut s'agir d'informations complémentaires qui indiquent à un opérateur de centre d'appels la personne à contacter lorsqu'une demande d'indemnisation est signalée pour un transfert à une unité spécialisée.

Les annotations peuvent être combinées à la valeur **Retourner (avec)** indiquée pour une dimension, pour fournir plus de détails. Tandis que la valeur **Retourner (avec)** est associée à la dimension ou à la décision en retour, l'annotation indique les règles à déclencher comme données d'entrée de la décision.

Utilisation de règles IBM Operational Decision Management

Les règles créées dans un système de gestion de règles métier (BRMS) tel qu'IBM Operational Decision Management (ODM) sont référencées et utilisées dans des applications IBM Analytical Decision Management. Ces règles externes peuvent être sélectionnées et utilisées de la même manière que d'autres règles en cliquant sur l'icône **Rechercher une règle existante** au point souhaité dans l'application. Les règles externes se distinguent par leur icône, mais s'utilisent de manière standard.

Pour obtenir des informations sur la création de règles externes à utiliser dans IBM Analytical Decision Management, voir *Application Designers Guide*.

La règle est insérée dans l'application de la même manière que n'importe quelle autre règle. Les sorties de règle doivent correspondre au contexte dans lequel la règle est utilisée ; par exemple, les affectations nécessitent une valeur (booléenne) true/false ou oui/non comme données d'entrée de la décision.

Comme pour les autres règles, cliquez sur le nom d'une règle externe utilisée dans l'application pour en voir les détails, notamment ses champs d'entrée, sa description et d'autres propriétés.

Si une règle externe comporte plusieurs champs de sortie, vous êtes invité à sélectionner le champ souhaité.

Si une règle externe est utilisée plutôt qu'un modèle, aucun champ cible n'est affiché, et toutes les sorties configurées sont répertoriées dans la liste déroulante.

Astuces pour l'utilisation des règles externes

- Les règles qui renvoient une valeur *true* ou *false* (ou *oui/non*, *0/1*) peuvent être utilisées dans toute sélection ou affectation comme données d'entrée d'une règle dans l'éditeur de règle, ou comme champ **Retourner (avec)** associé à un élément de dimension spécifique.
- Les règles qui renvoient des sorties catégorielles (telles que *rouge*, *blanc*, *bleu* ; *nord*, *sud*, *est*, *ouest*) peuvent être utilisées dans la section **Utiliser un modèle pour décider** de l'onglet Définir, comme données d'entrée de l'éditeur de règle ou comme champ **Retourner (avec)** associé à un élément de dimension.
- Les règles qui renvoient les sorties numériques peuvent être utilisées dans la section **Utiliser un modèle pour décider** de l'onglet Définir, comme données d'entrée d'une définition des priorités, comme données d'entrée dans l'éditeur de règle, ou comme champ **Retourner (avec)** associé à un élément de dimension.
- Les sources de données de vue de données analytiques prennent en charge l'utilisation des règles externes (hiérarchiques) complexes. Pour faire référence à une règle externe complexe, le modèle de données du projet doit être une source de *vue de données* et la fonctionnalité doit être activée (sur l'onglet Données, un administrateur doit sélectionner l'option appropriée **autorisation l'utilisation des règles ODM basées sur les données hiérarchiques**).
- Vous ne pouvez pas créer dans IBM Analytical Decision Management de règle partagée contenant une référence à une règle externe.

Chapitre 6. Génération de modèles prédictifs

A propos des modèles prédictifs

Les modèles permettent de prévoir ce qui est susceptible de se produire à l'avenir, d'après les motifs des données antérieures. Par exemple, vous pouvez utiliser un modèle pour prévoir les clients qui sont les moins susceptibles de résilier leur contrat ou les plus susceptibles de répondre à une offre particulière, en fonction de caractéristiques telles que le revenu, l'âge, ainsi que les organisations et les groupes auxquels ces clients sont abonnés.

Les modèles peuvent être utilisés de la même manière que les règles, mais alors que les règles peuvent être basées sur des politiques d'entreprise, la logique de l'entreprise ou sur d'autres hypothèses, les modèles sont générés à partir d'observations réelles de résultats antérieurs et peuvent découvrir des motifs qui, sinon, ne seraient pas apparents. Tandis que les règles dotent les applications d'une logique applicative commune, les modèles apportent des connaissances et un pouvoir prédictif.

Données pour générer et évaluer (scorer) des modèles

Deux différents types de données sont utilisés dans le processus de modélisation :

- Pour générer le modèle, vous avez besoin d'informations sur le point à prévoir. Par exemple, si vous voulez prévoir l'attrition, vous avez besoin d'informations sur les clients qui ont résilié leur contrat dans le passé. Ceci est souvent mentionné sous le nom de données d'historique ou de données d'analyse et doit en contenir tout ou partie des champs du modèle de données du projet, outre un champ supplémentaire qui enregistre le résultat à prévoir. Ce champ supplémentaire est utilisé comme **cible** pour la modélisation.
- Pour utiliser le modèle pour prédire des résultats ultérieurs, vous avez besoin de données sur le groupe ou la population qui vous intéresse (par exemple, les clients potentiels ou les demandes entrantes). Ceci est souvent mentionné sous le nom de données **opérationnelles** ou de données d'évaluation (scoring). Le modèle de données du projet est généralement basé sur ces données.
- Si le champ cible est inclus dans le modèle de données, la colonne **Opérationnel** ne doit pas être sélectionnée pour ce champ, car elle est indisponible lors de l'évaluation du modèle. Après la génération du modèle d'après des données d'historique, l'évaluation a pour objectif d'appliquer le modèle à de nouvelles données lorsque le résultat n'est pas déjà connu.

Génération de modèles

L'onglet Modélisation permet d'indiquer une source de données, un champ cible ainsi que d'autres options. Vous pouvez également sélectionner le type de modèle à générer.

Si des **paramètres** sont ajoutés à l'application (fichier de flux) dans IBM SPSS Modeler, un lien **Paramètres** est alors disponible sur les onglets Modélisation et Evaluation. Cliquez sur le lien **Paramètres** pour afficher les paramètres définis et leurs valeurs en cours. Vous pouvez également changer les valeurs de paramètre pour influencer sur la génération ou l'évaluation de modèle.

Génération d'un modèle prédictif

Lors de la génération d'un modèle prédictif, les seules spécifications obligatoires sont une source de données et un champ cible. Vous pouvez également sélectionner des options de préparation et de partitionnement des données, ainsi que de sélection des champs et des enregistrements à utiliser. Le modèle est généré en appliquant si nécessaire une gamme de méthodes au type cible et en identifiant automatiquement la technique ou la combinaison optimale pour les données sélectionnées.

Une fois évalué, le modèle renvoie un ou plusieurs champs contenant des prévisions, ainsi que les niveaux de fiabilité associés à ces prévisions. Par exemple, si le champ cible est un champ d'indicateur nommé *Response*, le modèle renvoie un champ nommé *XF-Response* contenant les valeurs prévues pour ce champ.

1. Indiquez une source de données. Il peut s'agir de toute source de données d'analyse enregistrant le résultat à prévoir.
2. Indiquez le champ cible dans lequel le résultat à prévoir sera enregistré. Par exemple, le champ cible peut indiquer les clients qui ont résilié leur contrat, répondu à des offres passées ou émis des demandes d'indemnisation frauduleuses.
Tous les autres champs sont automatiquement inclus comme données entrées pouvant être utiles pour prédire la valeur.
3. Indiquez des paramètres facultatifs si nécessaire. Pour plus d'informations, voir la rubrique «Paramètres facultatifs de modèle», à la page 65.
4. Si nécessaire, cliquez sur l'icône Présentation des données pour visualiser une présentation des données qui seront utilisées pour générer le modèle en cours. Pour plus d'informations, voir la rubrique «Présentation des données», à la page 45.
5. Cliquez sur **Générer un modèle**.
Vous pouvez fermer le navigateur ou travailler sur d'autres tâches pendant la génération du modèle. Une fois généré, le modèle est accessible dans la galerie. Pour plus d'informations, voir la rubrique «Galerie», à la page 28. Si la génération du modèle prend plus de temps que prévu, cliquez sur le bouton **Arrêter** pour revenir à la version précédemment enregistrée.
Lorsque la génération du modèle est terminée, les résultats s'affichent. Vous pouvez visualiser les **Résultats de modèles combinés** (graphiques de performances du modèle combiné généré) ou les **Résultats de modèles individuels** (détails de haut niveau des modèles individuels qui composent le modèle combiné).
Lorsque vous visualisez les *Résultats de modèles individuels*, le nom de chaque modèle est un lien qui ouvre une nouvelle boîte de dialogue contenant des informations sur le modèle (telles que le champ cible et les champs d'entrée). Si le modèle sélectionné est un *modèle d'arbre*, un visualiseur de fichiers s'affiche également.
6. Vous pouvez également employer les fonctions Evaluer et Test pour voir le fonctionnement du modèle avec vos données d'échantillon. Pour plus d'informations, voir la rubrique «Evaluation de modèles», à la page 69.

Résultats de la modélisation prédictive

Quand un modèle prédictif termine son exécution, la section de résultats vous permet de commuter entre l'affichage des **résultats des modèles combinés** et celui des **résultats des modèles pris séparément**. Les résultats des modèles combinés incluent des graphiques d'importance des variables et de fréquences pour le modèle combiné qui a été établi. Cela permet une évaluation visuelle rapide de la performance modèle et d'importance des différents prédicteurs pour obtention du résultat final. Les résultats des modèles pris séparément affichent des informations sur tous les modèles séparés constituant le modèle combiné. Vous pouvez ainsi cliquer sur chaque modèle pris séparément pour afficher ses détails.

Graphiques de fréquences

Les graphiques de fréquences des modèles affichent les valeurs observées sur l'axe horizontal, avec les valeurs prévues sur l'axe vertical ou une superposition. Ces représentations permettent de visualiser la fréquence des prévisions de chaque valeur ou de chaque réponse ainsi que la fréquence de l'exactitude de ces prévisions, fournissant ainsi une évaluation visuelle rapide de la précision du modèle.

- Pour les cibles d'indicateur ou d'ensemble, la distribution s'affiche dans un diagramme à barres, avec une barre pour chaque catégorie du champ cible. La hauteur de chaque barre indique le nombre de prévisions du résultat effectuées par le modèle. Les régions colorées dans chaque barre indiquent le nombre d'observations effectives de résultats concordants dans les données de test. Si le modèle est raisonnablement précis, le plus grand secteur dans chaque barre représente la réponse correspondante.

- Pour les cibles de plage, la distribution est représentée par un nuage de points en cellule. De nouveau, les valeurs prévues sont tracées sur l'axe horizontal et les valeurs observées, sur l'axe vertical. Idéalement, les points doivent être regroupés le long d'une diagonale. Plus les points (résultats) se dispersent et s'éloignent d'une distribution linéaire, moins le modèle est précis.

Graphiques d'importance des prédicteurs

En général, vous préférerez concentrer vos efforts de modélisation sur les champs prédicteurs les plus importants et abandonner ou ignorer les moins importants. Le graphique d'importance des prédicteurs peut vous y aider en indiquant l'importance relative de chaque prédicteur en estimant le modèle. Etant donné que les valeurs sont relatives, la somme des valeurs pour l'ensemble des prédicteurs affichée est 1,0. L'importance des prédicteurs n'a aucun rapport avec l'exactitude du modèle. Elle est juste rattachée à l'importance de chaque prédicteur pour la réalisation d'une prévision, peu importe si cette dernière est précise ou non.

Le calcul de l'importance des prédicteurs peut prendre un temps considérablement plus long que la construction de modèle, en particulier lors de l'utilisation de grands jeux de données.

Si l'option de partitionnement automatique des données a été sélectionnée lors de la génération du modèle, la partition de test sera automatiquement sélectionnée sur la page Evaluer.

Si le partitionnement n'a pas été activé, vous devez sélectionner une source de données pour évaluer le modèle. Il est recommandé de sélectionner une autre source de données que celle utilisée pour générer le modèle. Le partitionnement étant activé par défaut, ce paramétrage n'est normalement pas nécessaire sauf si vous l'avez changé.

Résultats de modèles individuels

Les résultats de modèles individuels affichent des informations sur chaque modèle individuel qui compose le modèle combiné. Le nom de chaque modèle est un lien qui ouvre une nouvelle boîte de dialogue contenant les détails complets de ce modèle, tel que le champ cible et les champs d'entrée. Si le modèle sélectionné est un *modèle d'arbre*, un visualiseur de fichiers s'affiche également. Ce visualiseur est interactif et fournit les fonctionnalités suivantes :

- Zoom avant et arrière sur l'arborescence
- Affichage des éléments d'arborescence comme tableaux, graphiques ou tableaux et graphiques combinés
- Changement de l'orientation de l'arborescence en descendant, de gauche à droite ou de droite à gauche
- Utilisation de la *fenêtre de carte d'arborescence* et des icônes en forme de flèche pour parcourir plus rapidement les arborescences volumineuses

Génération d'un modèle de règle prédictif

Les modèles de règle prédictifs permettent d'identifier les segments qui présentent une probabilité supérieure ou inférieure à un résultat déterminé. Par exemple, vous pouvez rechercher les clients associés à la probabilité de résiliation la plus faible ou les clients davantage susceptibles de répondre favorablement à une offre particulière. Le modèle obtenu ressemble à une liste de règles de sélection, avec un attribut **Inclure** ou **Exclure** qui indique le mode d'évaluation de chaque segment.

Les modèles de règle prédictifs s'apparentent aux règles de sélection, mais diffèrent par leur mode d'évaluation. Dans le cas d'une règle de sélection, les enregistrements exclus sont retirés, ce qui aboutit à un nombre inférieur d'enregistrements en sortie. Dans le cas d'un modèle, aucun enregistrement n'est retiré, mais des champs supplémentaires sont ajoutés pour indiquer des valeurs de prévision et de propension pour chaque enregistrement. Les segments inclus sont marqués comme *True* (ou associés à *False* ou à tout autre résultat que vous tentez de prévoir). Les segments exclus sont marqués comme *NULL*, de même que tous les enregistrements restants, car ils ne correspondent à aucune règle.

Chaque enregistrement est également associé à un score de propension qui indique la probabilité de la réponse cible, ce qui permet d'identifier les clients qui sont les plus susceptibles de répondre. Les modèles de règle prédictifs conviennent parfaitement pour générer des listes de diffusion et sont largement utilisés en gestion de la relation client, y compris pour les applications marketing ou de centre d'appels.

- La modélisation de règle prédictive n'est disponible que pour les cibles catégorielles.
- Vous pouvez visualiser, modifier et réorganiser les segments, sélectionner ceux à inclure ou à exclure à des fins d'évaluation. Vous pouvez par exemple choisir d'exclure un groupe de clients des offres à venir, en inclure d'autres et visualiser immédiatement la façon dont cette modification affecte votre taux de correspondance global.
- Comme pour les règles de sélection, vous pouvez ajouter ou éditer manuellement des segments si nécessaire. Il est souvent plus facile de laisser le modèle les trouver pour vous. Cependant, un certain nombre d'options sont proposées pour vous aider à affiner le résultat.
- Les règles sont appliquées dans l'ordre indiqué, la première règle correspondante déterminant le résultat pour un enregistrement donné. Prises individuellement, les règles ou les conditions peuvent se recouper, mais l'ordre des règles résout toute ambiguïté. Si aucune règle ne correspond, l'enregistrement est attribué à la règle restante.

Pour obtenir un modèle de règle prédictif

1. Indiquez une source de données. Il peut s'agir de toute source de données d'analyse enregistrant le résultat à prévoir.
2. Indiquez le champ cible dans lequel le résultat à prévoir sera enregistré.
3. Indiquez la valeur ou la réponse spécifique à trouver. Par exemple, si le champ cible est nommé *response*, vous pouvez sélectionner *True* ou *False* pour trouver les clients qui, respectivement, ont répondu ou n'ont pas répondu.
Par exemple, la cible peut indiquer les clients qui ont résilié leur contrat, répondu à des offres antérieures, émis des demandes d'indemnisation frauduleuses, etc. Tous les autres champs sont automatiquement inclus comme entrées pouvant être utiles pour prévoir la valeur.
4. Indiquez des paramètres facultatifs si nécessaire. Pour plus d'informations, voir la rubrique «Paramètres facultatifs de modèle», à la page 65.
5. Si nécessaire, cliquez sur l'icône Présentation des données pour visualiser une présentation des données qui seront utilisées pour générer le modèle en cours. Pour plus d'informations, voir la rubrique «Présentation des données», à la page 45.
6. Cliquez sur **Démarrer la génération**.
7. Indiquez si vous souhaitez trouver des segments à forte ou faible probabilité par rapport à la cible spécifiée. Par exemple, vous pouvez rechercher les groupes ou les clients les plus ou les moins susceptibles de résilier leur contrat, et les inclure en conséquence.
8. Indiquez le nombre maximal de segments à trouver. En général, ce nombre est maintenu petit à des fins de simplicité et de performances.
9. Cliquez sur **Développer un modèle** pour obtenir la règle de segment ou la liste de règles de segment qui identifie les groupes avec une probabilité supérieure ou inférieure.
10. Cliquez sur les icônes **Inclure/Exclure** pour indiquer le mode d'utilisation des règles ainsi que le mode d'évaluation des segments appropriés.
Les segments inclus sont dotés d'un score correspondant à la valeur cible (par exemple *response=true*). Les segments exclus sont marqués comme *NULL* mais ne sont pas retirés du fichier.
11. Vous pouvez ajouter ou éditer manuellement des segments si nécessaire. Pour plus d'informations, voir la rubrique «Définition de règles de segment», à la page 53.
12. Vous pouvez également employer les fonctions Evaluer et Test pour voir le fonctionnement du modèle avec vos données d'échantillon.

Conseils :

- Utilisez les flèches haut/bas pour déplacer les règles vers le haut ou vers le bas. Chaque enregistrement est évalué selon la première concordance trouvée ; par conséquent, l'ordre des règles change le résultat. Par exemple, un enregistrement correspondant à plusieurs règles est évalué d'après la première règle applicable.
- Changez une ou plusieurs options et re Cliquez sur **Développer un modèle** pour trouver d'autres règles. Par exemple, vous pouvez abaisser le seuil de segment ou d'intervalle de confiance minimum pour identifier des segments supplémentaires.
- Un clic sur **Développer un modèle** n'efface pas la table ni ne régénère le modèle à partir de zéro. Vous trouvez uniquement davantage de règles de segment à ajouter à la table.

Génération d'un modèle de groupement

Les modèles de groupement (ou modèles de classification) permettent de ranger les enregistrements dans un certain nombre de groupes (ou classes). Cette fonction peut vous aider à regrouper naturellement vos données.

Les modèles de groupement se chargent essentiellement d'identifier des groupes d'enregistrements similaires et de répertorier les enregistrements en fonction du groupe auquel ils appartiennent. Cette opération est effectuée sans l'aide des connaissances existantes sur les groupes et leurs caractéristiques. A dire vrai, vous ne connaîtrez peut-être même pas le nombre de groupes à rechercher. Ceci permet de distinguer les modèles de groupement des autres techniques d'apprentissage automatique. En outre, il n'existe aucun champ de sortie ou cible prédéfini pour le modèle à prévoir. Ces modèles sont souvent désignés "modèles d'**apprentissage non supervisé**" du fait de l'absence de norme externe permettant d'évaluer les performances de classification du modèle. Il n'y a aucune réponse *vraie* ni *fausse* pour ces modèles. Leur valeur est déterminée par leur capacité à capturer des groupements intéressants dans les données et ils fournissent des descriptions utiles de ces mêmes groupements.

Les méthodes de groupement reposent sur la mesure de la distance entre les enregistrements et les classes. Les enregistrements sont affectés à des classes de façon à réduire au maximum la distance entre les enregistrements appartenant à la même classe.

Les modèles de groupement sont souvent utilisés pour créer des classes ou des segments employés ensuite en tant qu'entrées dans les analyses ultérieures. Un exemple souvent utilisé est celui des segments de marché utilisés par les spécialistes du marketing pour partitionner leur marché global en sous-groupes homogènes. Chaque segment dispose de caractéristiques spéciales qui influent sur le succès des efforts de marketing entrepris pour ce segment. Si vous utilisez l'exploration de données pour optimiser votre stratégie marketing, vous pouvez généralement améliorer votre modèle de façon significative en identifiant les segments appropriés et en utilisant les informations des segments dans vos modèles prédictifs.

Pour obtenir un modèle de groupement

1. Indiquez une source de données. Il peut s'agir de toute source de données d'analyse enregistrant le résultat à prévoir.
2. Indiquez des paramètres facultatifs si nécessaire. Pour plus d'informations, voir la rubrique «Paramètres facultatifs de modèle», à la page 65.
3. Si nécessaire, cliquez sur l'icône Présentation des données pour visualiser une présentation des données qui seront utilisées pour générer le modèle en cours. Pour plus d'informations, voir la rubrique «Présentation des données», à la page 45.
4. Cliquez sur **Rechercher les classes**.
5. Facultativement, vous pouvez ajouter des classes manuelles. Pour plus d'informations, voir la rubrique «Utilisation de classes manuelles», à la page 64.
6. Vous pouvez également employer les fonctions Evaluer et Test pour voir le fonctionnement du modèle avec vos données d'échantillon.

Utilisation de classes manuelles

Vous pouvez définir manuellement des classes (clusters) puis, à l'aide du groupement automatisé, rechercher les classes d'enregistrements qui ne sont identifiées par aucune classe manuellement définie. Créez des classes manuelles à l'aide d'un jeu de règles "première correspondance" et d'affectations basées sur les segments. Les règles que vous créez définissent des classes et les autres servent pour la modélisation automatique de classe.

Les classes manuelles ne sont pas incluses dans les graphiques récapitulatifs des résultats.

1. Dans la section **Classe manuelle**, cliquez sur l'icône **Créer une règle** et ajoutez une ou plusieurs règles pour définir les classes manuelles.
2. Si vous le souhaitez, changez les noms de libellé de classe.
3. Cliquez sur l'icône **Enregistrements dans les classes manuelles** pour compter le nombre total d'enregistrements inclus dans les classes manuelles. La colonne **Effectif** est mise à jour pour chaque classe manuelle définie.

Génération d'un modèle d'association

Les modèles d'association permettent de prévoir les éléments davantage susceptibles d'apparaître ensemble ainsi que la puissance de leurs interrelations.

Les **règles d'association** associent une conclusion particulière (l'achat d'un produit particulier, par exemple) à un ensemble de conditions (l'achat de plusieurs autres produits, par exemple). Par exemple, la règle

bière <= conserves légumes & surgelés (173, 17,0 %, 0,84)

sous-entend que l'événement *bière* se produit lorsque les événements *conserves légumes* et *surgelés* ont lieu en même temps. La règle est fiable à 84 % et s'applique à 17 % des données ou à 173 enregistrements. Les algorithmes des règles d'association recherchent automatiquement les associations que vous pouvez trouver manuellement à l'aide de techniques de visualisation, telles que le noeud relations.

L'avantage des algorithmes de règles d'association par rapport aux algorithmes d'arbre décision standard (C5.0 et Arbres C&RT) est le fait qu'il puisse exister des associations entre *tous* les attributs. Un algorithme d'arbre décision peut construire une règle uniquement avec une seule conclusion. En revanche, les algorithmes d'association tentent d'en trouver plusieurs, chaque règle pouvant avoir une conclusion différente.

L'inconvénient des algorithmes d'association est qu'ils recherchent des motifs dans un espace de recherche très grand, nécessitant ainsi beaucoup plus de temps qu'un algorithme d'arbre décision. Les algorithmes utilisent une méthode de type **générer et tester** pour la recherche de règles. Tout d'abord, des règles simples sont générées, puis elles sont validées par rapport au jeu de données. Les règles satisfaisantes sont conservées et toutes les règles, soumises à différentes contraintes, sont alors spécialisées. La **spécialisation** consiste à ajouter des conditions à une règle. Ces nouvelles règles sont confrontées aux données pour être validées et lors de ce processus, les meilleures règles, ou les plus intéressantes, sont systématiquement conservées. L'utilisateur apporte généralement certaines limites au nombre possible d'antécédents pour savoir si une règle doit être conservée ou pas. Plusieurs techniques reposant sur la théorie de l'information ou des plans d'indexation efficaces sont utilisées pour réduire significativement l'espace de recherche.

Pour obtenir un modèle d'association

1. Sélectionnez une source de données. Il peut s'agir de toute source de données d'analyse enregistrant le résultat à prévoir.
2. Indiquez un format de données. Pour plus d'informations, voir la rubrique «Données tabulaires et données transactionnelles», à la page 65.
3. Indiquez des paramètres facultatifs si nécessaire. Pour plus d'informations, voir la rubrique «Paramètres facultatifs de modèle», à la page 65.

4. Si nécessaire, cliquez sur l'icône Présentation des données pour visualiser une présentation des données qui seront utilisées pour générer le modèle en cours. Pour plus d'informations, voir la rubrique «Présentation des données», à la page 45.
5. Cliquez sur **Générer un modèle**.
6. Vous pouvez également employer les fonctions Test pour voir le fonctionnement du modèle avec vos données d'échantillon. L'évaluation est indispensable pour les modèles de type association.

Avec d'autres types de modèle, vous pouvez sélectionner plusieurs techniques de modélisation. Pour des modèles d'association, le type de modèle *Apriori* est utilisé.

Données tabulaires et données transactionnelles

Lors de la génération d'un modèle d'association, des formats de données **tabulaires** ou **transactionnelles** sont disponibles. Le format de données tabulaires est sélectionné par défaut. Si vous choisissez **Transactionnelles**, sélectionnez des champs pour **ID** et **Contenu** dans les menus déroulants qui s'affichent.

Format transactionnel

Les données transactionnelles comportent un enregistrement distinct pour chaque transaction ou élément. Par exemple, si un client effectue plusieurs achats, chacun fait l'objet d'un enregistrement distinct, avec les articles associés liés par un ID client. Ce format est également parfois appelé format **till-roll** (déroulant).

Client	Achat
1	confiture
2	lait
3	confiture
3	pain
4	confiture
4	pain
4	lait

Données tabulaires

Les données tabulaires (également appelées données de **panier d'achat** ou données de **table des valeurs vraies**) comportent des éléments représentés par des indicateurs distincts, chaque champ indicateur indiquant la présence ou l'absence d'un élément spécifique. Chaque enregistrement représente un ensemble complet d'éléments associés. Les champs indicateurs peuvent être catégoriels ou numériques (certains modèles peuvent toutefois présenter des exigences plus particulières).

Client	Confiture	Pain	Lait
1	T	F	F
2	F	F	T
3	T	T	F
4	T	T	T

Paramètres facultatifs de modèle

Lors de la génération de modèles, vous pouvez sélectionner les paramètres facultatifs suivants. Selon le type de modèle et la cible sélectionnée, certains paramètres peuvent ne pas être visibles. Si l'un de ces paramètres est changé pour un modèle existant, celui-ci doit être régénéré pour que les changements s'appliquent.

Nettoyer et préparer automatiquement les données pour la génération de modèle fiable. Identifie et répare les problèmes de données pour rendre la modélisation plus rapide, plus prévisible et plus fiable. Examine les champs problématiques ou superflus, par exemple en manipulant les valeurs manquantes et extrêmes, en dérivant de nouveaux attributs si nécessaire et en améliorant les performances par des techniques intelligentes d'examen préalable et d'échantillonnage. Lors de la première génération d'un modèle avec une nouvelle source de données, l'analyse est effectuée pour identifier les problèmes et les correctifs, ce qui peut ralentir les performances sur la première passe seulement. Lors des exécutions suivantes, les correctifs sont appliqués, mais l'analyse n'est pas répétée sauf si la source de données change. Ce paramètre peut être désactivé pour certains modèles, notamment ceux dotés de paramètres personnalisés de préparation de données spécifiés par un utilisateur expérimenté.

Remarque : Les paramètres de préparation des données de mémoire automatique s'appliquent uniquement lors de la création d'un nouveau modèle et la mise au rebut n'est pas effectué.

Partitionner automatiquement les données pour activer le scoring du modèle dans la source de données de génération pour le scoring et le test. La sélection de cette option divise les données en sous-ensembles ou échantillons distincts à des fins d'apprentissage et de test du modèle. En générant le modèle sur un sous-ensemble et en le testant sur un autre, vous pouvez avoir une idée de la façon dont il va se généraliser à d'autres ensembles de données. Vous pouvez également indiquer un pourcentage pour la *Taille de la partition d'apprentissage (%)* et la *Taille de la partition de test*.

La taille minimale de la partition d'apprentissage est 1 et la taille maximale, 100. La taille minimale de la partition de test est 0 et la taille maximale, 100. La taille minimale de la partition de validation est 0 et la taille maximale, 100. Le pourcentage total doit être inférieur ou égal à 100 pour que la génération du modèle aboutisse.

Ces options de partitionnement sont disponibles pour tous les types de modèle excepté les modèles d'association.

Sélection des techniques de modélisation pour la génération de modèle. Si vous le souhaitez, cliquez sur **Sélectionner** pour sélectionner les techniques de modélisation à inclure lors de la génération du modèle. En fonction de la cible sélectionnée, les techniques de modélisation disponibles sont répertoriées. Les techniques de modélisation disponibles (et celles sélectionnées par défaut) varient selon la cible sélectionnée. Par exemple, l'algorithme *Liste de décisions* n'apparaît que pour les cibles *indicateur*. Cette fonction n'est disponible que pour les modèles prédictifs.

Pour obtenir des détails complets sur les techniques de modélisation, voir *IBM SPSS Modeler Algorithms Guide* ainsi que la documentation fournie avec le produit IBM SPSS Modeler et disponible sur le Web.

Nombre maximum de modèles à combiner. Permet de définir le nombre maximal de modèles à conserver et à combiner. Si vous définissez cette option sur 1, un seul modèle est généré et conservé. Si le nombre de modèles générés est inférieur à la valeur entrée ici, alors tous les modèles générés sont conservés et combinés. Cette option n'est disponible que pour les modèles prédictifs.

Critères de profit (permettent de trier les modèles avec des cibles binaires). Permet de définir des critères de profit avant de générer le modèle. Les valeurs définies ici sont définies comme valeur par défaut pour l'évaluation du modèle. Cette option n'est disponible que pour les modèles prédictifs pour lesquels une cible indicateur est sélectionnée.

Indiquer les entrées à utiliser. Permet de sélectionner les champs à utiliser. En général, il s'agit de champs ayant une relation pratique avec l'élément que vous essayez de prévoir (par exemple, l'âge ou le revenu). Si vous disposez d'un grand ensemble de données, la limitation du nombre de champs constitue l'une des manières de simplifier le modèle. Lorsque les données contiennent des champs tels que l'identificateur de client ou le numéro de contact, elles ne sont généralement pas utiles en modélisation et ne doivent pas être sélectionnées. Les champs qui dupliquent d'autres données peuvent également être exclus.

Lorsque vous cliquez sur un champ d'entrée lié (une expression), vous ouvrez l'afficheur d'expression pour l'expression concernée. Pour éditer une expression, voir l'onglet Données. Pour plus d'informations, voir la rubrique «Générateur de formules», à la page 49.

Indiquer les sélections à utiliser. Indique les enregistrements à inclure ou à exclure lors de la modélisation. Vous pouvez rechercher des règles existantes ou en créer des nouvelles, selon le cas. Pour plus d'informations, voir la rubrique «Définition de règles de sélection», à la page 55.

En outre, si des sélections globales ont été définies, elles s'affichent ici, et vous pouvez indiquer si elles doivent également s'appliquer pendant la modélisation. Vous pouvez sélectionner toutes les règles de sélection globale ou n'en indiquer aucune ; vous ne pouvez pas sélectionner de sous-ensemble.

Options de segmentation

Lorsque vous générez des modèles de règle prédictifs, un certain nombre d'options de segmentation vous sont proposées, en complément des options disponibles avec d'autres types de modèle.

Taille de segment minimale. Ces paramètres déterminent la taille de segment minimale, la plus grande des deux valeurs étant prioritaire. Par exemple, si la valeur de pourcentage est supérieure à la valeur absolue, la valeur de pourcentage est prioritaire.

- **Pourcentage du segment précédent.** Spécifie la taille minimale du groupe en pourcentage d'enregistrements. Les valeurs minimale et maximale autorisées sont respectivement 0 et 99,9.
- **Valeur absolue.** Spécifie la taille minimale du groupe en nombre absolu d'enregistrements. La valeur minimale autorisée est 1. Aucune valeur maximale n'est définie.

Nombre maximal d'attributs. Indique le nombre maximal de conditions par règle de segment. La valeur minimale autorisée est 1. Aucune valeur maximale n'est définie.

Autoriser la réutilisation d'attribut. Si cette option est activée, chaque cycle peut prendre en compte tous les attributs, même ceux qui ont été utilisés dans des cycles précédents. Les conditions d'un segment se font sous forme de cycles, où chaque cycle ajoute une nouvelle condition. Le nombre de cycles est défini à l'aide du paramètre **Nombre maximal d'attributs**.

Intervalle de confiance pour les nouvelles conditions. Spécifie le niveau de confiance pour les tests de signification des segments. Ce paramètre joue un rôle significatif dans le nombre de segments (le cas échéant) renvoyés ainsi que dans le nombre de conditions par règle de segment. Plus la valeur est élevée, plus l'ensemble de résultats renvoyé est petit. Les valeurs minimale et maximale autorisées sont respectivement 50 et 99,9.

Options de classe auto

Lorsque vous générez des modèles de groupement, un certain nombre d'options de classe automatique vous sont proposées, en complément des options disponibles avec d'autres types de modèle.

Champ d'évaluation. Indique le champ à utiliser pour l'évaluation. Si un champ d'évaluation est sélectionné, la méthode de classement dans le noeud de génération de classe auto d'IBM SPSS Modeler est définie sur *Importance de l'évaluation*. Si aucun champ d'évaluation n'est sélectionné, la méthode de classement est laissée sur sa valeur par défaut. Dans ce cas, il est possible de sélectionner un champ analytique comme champ d'évaluation.

Définir une plage pour le nombre de classes à trouver. Indique le nombre minimal et le nombre maximal de classes que le modèle de classe auto doit trouver.

Options d'évaluation et de règle d'association

Lorsque vous générez des modèles d'association, un certain nombre d'options d'évaluation et de règle d'association vous sont proposées, en complément de quelques options disponibles avec d'autres types de modèle.

Options de règle d'association.

- **Couverture (%)**. Indiquez le pourcentage de couverture minimum permettant de contrôler l'applicabilité générale des règles. Par exemple, une valeur définie sur 80 % signifie que toutes les règles s'appliqueront à au moins 80 % des observations. Une valeur plus importante génère moins de règles ou potentiellement aucune règle.
- **Confiance (%)**. Indiquez le pourcentage de confiance minimum.
- **Nombre maximal de conditions préalables**. Indiquez le nombre maximal de conditions préalables.
- **Les données sont déjà triées par ID (format transactionnel uniquement)**. Sélectionnez cette option pour désactiver le tri par défaut au début de la génération de modèle permettant de garantir la contiguïté des ID. Sélectionner cette option réduit la durée de traitement. Cette option n'a aucun impact si le format de données est tabulaire.

Options d'évaluation.

- **Nombre maximal de prévisions**. Indiquez le nombre maximal de prévisions incluses. Ce paramètre est utilisé avec la liste déroulante Critère de règle pour produire les meilleures prévisions, c'est-à-dire les prévisions qui présentent, par exemple, le plus haut niveau de fiabilité, de support et de lift.
- **Critère de règle**. Sélectionnez la mesure utilisée pour déterminer la puissance des règles. Les règles sont triées selon la force des critères sélectionnés ici afin de renvoyer les prévisions de niveau supérieur d'un jeu d'éléments. Les critères disponibles sont les suivants :
 - **Confiance** affiche le rapport entre le support de règle et le support de condition de règle. Cette option indique la proportion d'enregistrements correspondant à la condition ou aux conditions de règle spécifiées, pour laquelle la prévision ou les prévisions de règle sont également vraies. Par exemple, si 50 % des données d'apprentissage contiennent "pain" (support de condition de règle) mais que seulement 20 % contiennent à la fois "pain" et "fromage" (support de règle), alors la prévision de la règle pain -> fromage sera Support de règle / Support de condition de règle, soit 40 %.
 - **Couverture (%)** affiche le support de condition de règle, c'est-à-dire la proportion d'identificateurs pour laquelle les conditions de règle sont vraies, d'après les données d'apprentissage. Par exemple, si 50 % des données d'apprentissage incluent l'achat de pain, alors le support de condition de la règle pain -> fromage est égal à 50 %.
 - **Support de règle (%)** affiche la proportion d'enregistrements pour laquelle la règle, la ou les conditions de règle et la ou les prévisions de règle sont vraies. Par exemple, si 20 % des données d'apprentissage contiennent "pain" et "fromage", le support de la règle pain > fromage est de 20 %.
 - **Lift** affiche le rapport entre le niveau de fiabilité de la règle et la probabilité antérieure de disposer de la prévision de règle. Par exemple, si 10 % de la population entière achète du pain, alors une règle qui prévoit si les personnes achèteront du pain avec 20 % de confiance aura un lift de $20/10 = 2$. Si une autre règle vous dit que les personnes achèteront du pain avec une confiance de 11 % alors la règle a un lift proche de 1, ce qui signifie qu'avoir la/les condition(s) de règle ne fait pas beaucoup de différence quant à la probabilité d'avoir la prédiction de règle. En général, les règles dont le lift n'est pas 1 sont plus intéressantes que celles dont le lift est proche de 1.
 - **Capacité de déploiement (%)** mesure le pourcentage de données d'apprentissage qui répond à la ou aux conditions de règle mais pas à la prévision de règle. En termes d'acquisition de produits, cette option indique essentiellement le pourcentage de la base totale de clients qui possède (ou a acheté) la ou les conditions de règle mais qui n'a pas encore acheté la ou les prévisions de règle. Les statistiques de déployabilité sont définies comme suit : $((\text{Support de condition de règle dans nombre d'enregistrements} - \text{Support de règle dans nombre d'enregistrements}) * 100)$, où Support de condition de règle représente le nombre d'enregistrements pour lequel les conditions de règle sont vraies et Support de règle, le nombre d'enregistrements pour lequel les conditions de règle et la prévision sont vraies.
- **Vérifier toute la transaction**. Sélectionnez des options parmi les suivantes :
 - **S'assurer qu'il n'y a pas de prévisions** garantit que les prévisions de règle ne sont pas présentes. Par exemple, si l'objectif du scoring est d'effectuer une recommandation de meubles pour la maison, il est alors peu probable qu'un panier qui contient déjà une table pour dîner en achètera une autre.

Dans ce cas, sélectionnez cette option. Cependant, si les produits concernés sont périssables ou jetables (comme le fromage, le lait pour nourrisson ou les mouchoirs), les règles dans lesquelles la conséquence figure déjà dans le panier peuvent être valables. Dans ce dernier cas, l'option la plus utile peut être **Ne pas vérifier les prévisions**.

- **S'assurer qu'il y a des prévisions** garantit que les prévisions de règle sont également présentes dans le panier. Cette approche est utile lorsque vous essayez de vous faire une idée des transactions ou des clients existants. Par exemple, vous souhaitez peut-être identifier les règles ayant le plus grand lift, puis explorer les clients qui respectent ces règles.
- **Ne pas vérifier les prévisions** inclut toutes les règles lors d'une évaluation, que des prévisions de règle soient présentes ou absentes du panier.
- **Autoriser la répétition des prévisions.** Sélectionnez cette option pour inclure plusieurs règles avec la même prévision de règle lors de l'évaluation. Par exemple, sélectionner cette option permet d'évaluer les règles suivantes :
pain & fromage -> vin
fromage & fruit -> vin
Désactivez cette option pour exclure la répétition des prévisions lors de l'évaluation.
- **Ignorer les éléments sans correspondance.** Sélectionnez cette option pour ignorer la présence d'éléments supplémentaires dans l'ensemble d'éléments. Par exemple, lorsque cette option est sélectionnée pour un panier qui contient [tente & sac de couchage & bouilloire], la règle tente & sac de couchage -> gazinière s'applique malgré l'élément supplémentaire (bouilloire) présent dans le panier.

Evaluation de modèles

Lorsque vous avez généré ou ouvert un modèle terminé, cliquez sur **Evaluer** pour obtenir des graphiques (par exemple, de fréquences, de fréquences des erreurs, de matrice de coïncidence, de gain, de lift, de réponse, de profits et de retour sur investissement) ainsi que pour simuler le profit maximal possible. Les types de graphique et les autres options disponibles varient selon le type de modèle sélectionné. La boîte de dialogue Evaluer n'est disponible que si le modèle sélectionné est terminé. L'évaluation est indisponible pour les modèles de type association.

1. Dans le panneau supérieur, sélectionnez un ensemble de données d'évaluation. Comme pour la génération de modèle, il doit s'agir d'un ensemble de données d'analyse pour lequel des réponses réelles sont connues. (Sinon, vous n'avez aucun moyen de comparer les réponses réelles à celles prévues par le modèle.)
2. Pour modifier les modèles à inclure dans l'évaluation, cliquez sur **Modifier la sélection de modèle**. Lorsque vous avez effectué vos sélections, vous pouvez également cliquer sur le bouton suivant **Copier les sélections de modèle en cours dans la table de résultats du modèle individuel**. Ces options ne sont disponibles que pour les modèles de type prédictif.



Figure 13. Copier les sélections de modèle en cours dans la table de résultats du modèle individuel

3. Sélectionnez les graphiques à générer.
4. Si vous simulez le profit, sélectionnez la case **Simuler le profit maximum** et entrez la population sur laquelle baser le calcul du profit, le budget et la valeur minimale. Pour plus d'informations, voir la rubrique «Simulation du profit maximal», à la page 71.
5. Si vous demandez un graphique de profits ou de retour sur investissement ou que vous simulez le profit, précisez les valeurs de coût et de revenu. Ces valeurs indiquent, respectivement, le coût de chaque offre et le montant du revenu escompté à partir de chaque réponse favorable.
6. Sélectionnez la valeur de réponse considérée comme concordante. Il doit s'agir de l'une des valeurs du champ cible utilisé pour générer le modèle ; par exemple, *response = true* pour indiquer une réponse positive ou favorable.

7. Si vous demandez un graphique de performances, vous pouvez également sélectionner l'option **Inclure les évaluations de modèle individuelles pour les graphiques de performances**.
8. Si vous le souhaitez, sélectionnez l'option d'**inclusion des résultats sous forme de tableau**. Dans ce cas, le retour des résultats de l'évaluation prend davantage de temps.
9. Cliquez sur **Exécution** pour visualiser les résultats.

Graphiques de fréquences (toutes les cibles)

Les graphiques de fréquences représentent les valeurs observées par rapport aux valeurs attendues et s'affichent par défaut lorsque le modèle est terminé. La boîte de dialogue Evaluer permet d'obtenir des graphiques de fréquences pour un autre ensemble de données que celui de test. Pour les cibles de plage, le graphique de fréquences est le seul type de graphique disponible.

Vous pouvez sélectionner l'un des graphiques supplémentaires ci-après.

Graphiques de fréquences des erreurs

Les graphiques de fréquences des erreurs affichent un histogramme des erreurs dans les résultats du modèle, l'erreur étant la différence entre la valeur prévue et la valeur réelle. Une erreur positive ou négative se produit lorsque la valeur prévue est respectivement supérieure ou inférieure à la valeur réelle. Ce graphique n'est disponible que pour les cibles continues associées au type de modèle prédictif.

Graphiques de matrice de coïncidence

Une matrice de coïncidence représente le nombre de résultats *prévus* par rapport aux résultats *réels*. Pour une cible numérique, la plage est divisée en nombre de casiers indiqués.

Graphiques de gain

Les gains sont utilisés pour afficher le gain ou "lift" pouvant être obtenu à l'aide du modèle. Ils sont définis comme la proportion de concordances dans chaque incrément par rapport au nombre total de concordances disponibles. La diagonale plate trace la réponse attendue pour tout l'échantillon en l'absence de modèle. Dans ce cas, le taux de réponse serait constant, car une personne est susceptible de répondre comme une autre. Pour doubler ce nombre, vous devriez interroger deux fois plus de personnes. La courbe indique la quantité d'amélioration possible du résultat si vous ciblez les personnes les plus susceptibles de répondre. Plus la courbe s'accroît, plus les gains sont élevés.

Graphiques de lift

Le graphique de lift représente le pourcentage d'enregistrements qui sont des concordances dans chaque incrément ainsi que le pourcentage global de concordances dans l'ensemble de données d'apprentissage.

Graphiques de réponse

Le graphique de réponse trace le pourcentage d'enregistrements qui sont des concordances dans l'incrément.

Graphiques de profits

Le profit est égal au revenu de chaque enregistrement moins le coût de l'enregistrement. Les profits d'un quantile correspondent à la somme des profits de tous ses enregistrements. Les revenus sont supposés ne s'appliquer qu'aux concordances, tandis que les coûts concernent tous les enregistrements. Pour plus d'informations, voir la rubrique «Simulation du profit maximal», à la page 71.

Graphiques de retour sur investissement

Le retour sur investissement (ROI, Return On Investment) s'apparente au profit car il implique la définition de revenus et de coûts. Le graphique de retour sur investissement compare les profits aux coûts pour chaque intervalle.

Simulation du profit maximal

Sur la base des taux de réponse prévus par le modèle, vous pouvez simuler le profit maximal à faire pour une population donnée. A cet effet, vous devez indiquer les valeurs de coût et de revenu, ainsi que la taille de la population.

Pour prévoir le profit maximal :

1. Indiquez la population (nombre total d'enregistrements) sur laquelle la prévision doit être basée.
Remarque : lors de l'évaluation d'un modèle, cette valeur est automatiquement définie sur le nombre total d'enregistrements dans la source de données ou le sous-ensemble.
2. Indiquez une valeur de budget si vous souhaitez que le paramètre de budget soit inclus dans la simulation de profit, et une valeur minimale.
3. Précisez le coût associé à chaque enregistrement. Par exemple, il peut s'agir du coût d'impression et d'envoi, par personne, de la documentation promotionnelle pour une campagne.
4. Indiquez le revenu attendu d'une réponse positive. Par exemple, le coût au détail d'un élément dans une campagne.
5. Sélectionnez la réponse qui vous intéresse.
6. Cliquez sur **Exécuter**.

Le graphique de résultats prévu montre le meilleur résultat pour le profit maximal et identifie le pourcentage d'enregistrements que vous devez cibler pour y parvenir, selon la propension à répondre identifiée par le modèle.

Sous le graphique, vous pouvez déplacer un curseur gradué de zéro à 100 % des enregistrements à cibler, pour voir les incidences correspondantes sur le profit.

Lorsque les chiffres sont acceptables, vous pouvez utiliser les paramètres pour évaluer les enregistrements. A cet effet, cliquez sur **Utiliser pour le scoring**. Le pourcentage d'enregistrements à cibler s'affiche dans le champ du **meilleur pourcentage**, dans la section Options de propension de l'onglet Evaluation.

Comparaison de profits

Pour vérifier que les enregistrements sélectionnés sont pertinents, vous pouvez comparer le profit probable attendu avec ces enregistrements, un groupe aléatoire du même pourcentage d'enregistrements et l'ensemble complet des enregistrements.

Les détails présentés incluent le nombre d'enregistrements ciblés, le taux probable de réponse, les coûts, le revenu et le profit global. Vous pouvez exporter ces informations au format HTML pour les utiliser dans d'autres applications.

Graphique de profits

Vous pouvez afficher le profit cumulatif ou non cumulatif attendu des enregistrements sélectionnés sous la forme d'un graphique.

Remarque : ce graphique est généré lors de la première exécution de l'évaluation et n'est pas automatiquement mis à jour avec les changements que vous apportez avec le curseur dans la section de la simulation de profit.

Test du modèle

La boîte de dialogue Test permet de voir comment un ou plusieurs exemples d'enregistrements seront évalués. Les résultats sont propres au contexte dans lequel le test est exécuté et peuvent refléter les scores renvoyés par un modèle ou une règle. Voici quelques exemples des types de résultats que vous pouvez visualiser selon votre type d'application :

- des clients spécifiques et voir les offres affectées à chacun d'eux ;
 - des enregistrements de client et voir les clients susceptibles de répondre à une campagne de publipostage ;
 - des demandes d'indemnisation spécifiques et voir les actions affectées à chacune d'elles ;
 - des cas particuliers et voir leur évaluation pour chaque règle.
1. Sur l'onglet Modélisation, cliquez sur **Test**. La boîte de dialogue Test apparaît.
 2. Sélectionnez la source de données à utiliser ou sélectionnez **Indiquer des données personnalisées** dans la section Enregistrements pour entrer les enregistrements de test manuellement.
 3. Indiquez les sélections à utiliser, le cas échéant. Par exemple, vous pouvez inclure ou exclure certains enregistrements du test. Vous pouvez rechercher des règles existantes ou en créer des nouvelles, selon le cas. Pour plus d'informations, voir la rubrique «Définition de règles de sélection», à la page 55.
 4. Si vous utilisez une source de données, indiquez le nombre d'enregistrements que vous voulez renvoyer. Vous pouvez renvoyer les *N* premiers enregistrements jusqu'à 100 au maximum, lus à partir du sommet des données ou d'un échantillon aléatoire d'enregistrements. Si vous testez un modèle d'association avec des données transactionnelles, les données d'entrée sont échantillonnées au niveau des transactions, pas au niveau des enregistrements.
 5. Si vous utilisez des données personnalisées, cliquez sur **Indiquer les données personnalisées** et cliquez sur **Ajouter l'enregistrement personnalisé**. Tous les champs définis dans le modèle de données du projet sont énumérés, ce qui vous permet d'indiquer une valeur pour chacun d'eux. Cliquez sur **Enregistrer** lorsque vous avez terminé.
 6. Sous **Sélectionner les champs à afficher**, sélectionnez les champs qui doivent apparaître dans la sortie de test.
 7. Cliquez sur **Exécution**. Le récapitulatif de test affiche une liste d'affectations réussies pour chaque enregistrement.
 8. Pour obtenir les détails d'une ligne, cliquez sur l'icône d'information en regard de celle-ci.

Utilisation de scores de modèle dans des applications

Évaluer un modèle signifie l'appliquer à des données ou à une population en vue d'obtenir un résultat ou une prévision utilisable comme données d'entrée de décisions. Par exemple, vous pouvez évaluer tous les clients de votre base de données afin d'identifier les prospects pour une campagne de publipostage, ou évaluer chaque client qui contacte votre centre d'appels pour décider quels produits ou services lui offrir. Selon l'application, les résultats d'évaluation peuvent être écrits dans une table de base de données ou dans un fichier à plat, ou être utilisés comme données d'entrée du segment, de la sélection et des règles d'affectation qui régissent les décisions dans une application.

Les scores de modèle sont utilisables des manières suivantes :

- Comme données d'entrée de règles. Pour plus d'informations, voir la rubrique «Création des règles», à la page 53.
- Pour les applications qui prennent en charge l'évaluation par lots, les scores de modèle peuvent être écrits dans une table de base de données ou dans un fichier à plat. Pour plus d'informations, voir la rubrique «Evaluation (scoring) de modèles vers une table de base de données, un fichier, un serveur analytique ou un serveur IBM Cognos BI», à la page 75.

Champs de sortie du modèle

Par convention, les noms des champs de sortie d'évaluation sont basés sur le nom de champ cible, auquel est ajouté un préfixe personnalisé. Par exemple si le champ cible est appelé *Response*, les champs de sortie porteront des noms tels que *\$XF-Response* et *\$XFRP-Response*. Les champs spécifiques dépendent du type de modèle et du niveau de mesure appliqué au champ cible sélectionné, comme détaillé ci-après. Pour plus d'informations, voir la rubrique «Niveaux de mesure», à la page 46.

Cibles indicateur. Les modèles automatisés avec cibles indicateur affichent les champs suivants :

- **\$XF-<target>**. Valeur prévue pour chaque enregistrement, rapportée en termes de valeurs "true" et "false" définies pour le champ.
- **\$XFRP-<target>**. Score de propension ; indique la probabilité de la valeur "true" pour chaque enregistrement. Les scores de propension diffèrent des cotes de confiance, qui sont typiquement rapportées en termes de prévision pour l'enregistrement concerné. Par exemple, une prévision fausse ("false") dotée d'un niveau de fiabilité élevé se traduit par une propension basse, car elle indique une forte probabilité de ne *pas* répondre.

Cibles catégorielles. Les modèles automatisés avec cibles catégorielles (nominales ou ordinales) affichent les champs suivants :

- **\$XS-<target>**. Valeur prévue pour chaque enregistrement.
- **\$XSC-<target>**. Valeur de confiance associée à la prévision.

Cibles continues. Les modèles automatisés avec cibles numériques continues affichent les champs suivants :

- **\$XR-<target>**. Valeur prévue pour chaque enregistrement.
- **\$XRE-<target>**. Erreur standard pour la prévision.

Modèles de règle prédictifs (interactifs). Les modèles de règle prédictifs (cibles indicateur ou catégorielles uniquement) affichent les champs suivants :

- **\$D-<target>**. Résultat prévu pour chaque enregistrement ; la valeur "true" est définie pour le champ utilisé pour indiquer la réponse souhaitée et NULL représente toute autre valeur.
- **\$DP-<target>**. Probabilité de résultat souhaité pour chaque enregistrement ; correspond à la probabilité affichée pour le segment dans la table de résultats lors de la génération du modèle. Pour plus d'informations, voir la rubrique «Génération d'un modèle de règle prédictif», à la page 61.
- **\$DI-<target>**. Numéro d'index utilisé pour identifier le segment dans lequel l'enregistrement se situe (s'affiche dans la colonne la plus à gauche dans les résultats de modélisation).
- **\$DRP-<target>**. Score de propension indiquant la probabilité de la réponse souhaitée. Disponible pour les cibles indicateur uniquement.

Chapitre 7. Evaluation de modèles

Evaluation (scoring) de modèles vers une table de base de données, un fichier, un serveur analytique ou un serveur IBM Cognos BI

Évaluer un modèle signifie l'appliquer à l'échantillon de données ou à la population qui vous intéresse. Par exemple, vous pouvez utiliser vos clients existants pour générer un modèle qui prévoit les clients qui sont les plus susceptibles de répondre à une campagne. Ensuite, vous pouvez employer ce modèle pour évaluer des enregistrements relatifs à des personnes qui ne font pas encore partie de votre clientèle, mais que vous pouvez inclure dans un mailing promotionnel.

Les étapes de base de l'évaluation d'un modèle sont les suivantes :

1. Sélectionnez la source de données et les sous-ensembles de ces données pour évaluation.
2. Sélectionnez la destination des données évaluées.
3. Sélectionnez les informations de sortie et leur mode d'évaluation.
4. Indiquez les enregistrements à évaluer. Dans la liste déroulante Evaluation, indiquez l'option d'évaluation choisie : **Tous les enregistrements**, **Top N%**, **Top n enregistrements**, **Propension Min/Max** (disponible uniquement pour les cibles *indicateur*) ou **Score minimum** (disponible uniquement pour les cibles *continues*).
5. Prévisualisez les résultats attendus et exécutez une présentation des données, si nécessaire.
6. Cliquez sur **Évaluer maintenant** pour exécuter l'évaluation sur le modèle avec vos paramètres. Lorsque vous avez terminé, examinez la section Rapport récapitulatif d'évaluation. Pour plus d'informations, voir la rubrique «Rapport récapitulatif d'évaluation», à la page 82.

Les scores sont écrits dans une ou plusieurs zones ajoutées aux données. Par convention, les noms de ces zones sont basés sur le nom de zone cible, auquel est ajouté un préfixe. Par exemple, si la zone cible est appelée *Response*, les zones de sortie porteront des noms tels que *\$XF-Response* et *\$XFRP-Response*. Pour plus d'informations, voir la rubrique «Sélection de champs de sortie», à la page 81.

Remarque : Lorsque les données source proviennent d'une base de données, l'exécution est renvoyée par commande push à la base de données si possible (dans un processus connu sous le nom de "SQL push-back"). Généralement IBM Analytical Decision Management prend en charge le même niveau de SQL push-back qu'IBM SPSS Modeler mais avec les limitations suivantes :

- Si le projet inclut des zones ou des règles de sélection supplémentaires avec des références à des modèles ou à des règles externes, ces références ne sont pas renvoyées à la base de données par commande push. Cela peut diminuer le nombre d'actions réalisables par ce biais pour le flux de projet dans son ensemble.
- Dans les applications autres qu'IBM SPSS Modeler Advantage, la fonction de renvoi par commande push est activée uniquement pour les règles de sélection (sélections globales et sélections de membre pour la première dimension). Le degré de renvoi par commande push peut être accru si vous sélectionnez une zone entière comme ID d'entité dans le modèle de données du projet.

IBM SPSS Modeler Advantage

Pour **IBM SPSS Modeler Advantage**, le flux enregistré contiendra deux branches : la branche de création et la branche d'évaluation.

Le noeud source de la branche de **création** se basera toujours sur la source du modèle de données du projet. Il comprendra les zones marquées comme opérationnelles ou non opérationnelles.

Le noeud source de la branche d'**évaluation** se basera également toujours sur la source du modèle de données du projet. En revanche, il n'inclura que les zones marquées comme opérationnelles. Les zones marquées comme non opérationnelles ne seront pas incluses puisqu'elles ne sont pas nécessaires à l'évaluation. Par exemple, la zone cible d'un modèle prédictif est en général associée à un indicateur "non opérationnel" et n'est pas nécessaire en tant qu'entrée d'évaluation.

Par défaut, la branche d'évaluation d'un flux enregistré disposera d'un noeud terminal Table. Ce noeud servira à référencer la branche d'évaluation dans les paramètres de déploiement du flux. Le service d'évaluation d'IBM SPSS Collaboration and Deployment Services utilisera ces informations pour déterminer le noeud terminal à utiliser lors d'une évaluation en temps réel.

S'il est configuré, l'onglet **Evaluation** affichera une section pour l'**Evaluation par lot**. Sous **Evaluation par lot** se trouve la section **Options d'évaluation**. Ces paramètres ne s'appliquent que dans le cadre d'une évaluation au sein de l'interface utilisateur IBM Analytical Decision Management, en cliquant sur le bouton **Evaluer maintenant**. Ils ne seront pas reflétés dans le flux généré enregistré. Aussi, ils ne s'appliquent pas dans le cadre d'une évaluation via le client IBM SPSS Modeler, ni lors de la planification d'un travail via le client IBM SPSS Collaboration and Deployment Services Deployment Manager.

L'évaluation par lot peut également sélectionner les zones de destination et de sortie. Lors de l'enregistrement du flux, un noeud terminal supplémentaire sera créé sur la branche d'évaluation. Les paramètres de zone de destination et de sortie permettront de créer un noeud terminal Export sur la branche d'évaluation. Ces paramètres s'appliquent lors de l'exécution du flux depuis le client IBM SPSS Collaboration and Deployment Services Deployment Manager ou lors de la planification d'un travail via le client IBM SPSS Collaboration and Deployment Services Deployment Manager.

Projets non IBM SPSS Modeler Advantage

Pour les projets **autres qu'IBM SPSS Modeler Advantage**, le flux enregistré ne disposera que de la branche d'évaluation. Comme pour la branche d'évaluation d'IBM SPSS Modeler Advantage, le noeud source de la branche d'évaluation se basera toujours sur la source du modèle de données du projet. De même, il n'inclura que les zones marquées comme opérationnelles. Si des tables secondaires sont utilisées, des noeuds sources supplémentaires seront créés. Ces noeuds sources secondaires sont définie par les sources de données secondaires associées à la source de données principale.

L'onglet **Déployer**, lorsqu'il est configuré, peut présenter les sections **Evaluation en temps réel** et **Evaluation par lots**. Ces paramètres sont indépendants l'un de l'autre.

Lors de l'enregistrement du flux, les paramètres d'**Evaluation en temps réel** permettront de créer un noeud terminal Table sur la branche d'évaluation. Ce noeud table permettra de référencer la branche d'évaluation dans les paramètres de déploiement du flux enregistré. Le service d'évaluation d'IBM SPSS Collaboration and Deployment Services utilisera ces informations pour déterminer le noeud terminal à utiliser lors d'une évaluation en temps réel.

Sous **Evaluation par lot** se trouve la section **Options d'évaluation**. Ces paramètres ne s'appliquent que dans le cadre d'une évaluation au sein de l'interface utilisateur IBM Analytical Decision Management, en cliquant sur le bouton **Evaluer maintenant**. Ils ne seront pas reflétés dans le flux généré enregistré. Aussi, ils ne s'appliquent pas dans le cadre d'une évaluation via le service d'évaluation IBM SPSS Collaboration and Deployment Services, ni lors de la planification d'un travail via un client IBM SPSS Collaboration and Deployment Services Deployment Manager.

L'évaluation par lot peut également sélectionner les zones de destination et de sortie. Lors de l'enregistrement du flux, ces paramètres permettront de créer un noeud terminal Export sur la branche d'évaluation. Ces paramètres s'appliquent lors de la planification d'un travail via le client IBM SPSS Collaboration and Deployment Services Deployment Manager ou lors de l'exécution d'un flux du client IBM SPSS Modeler.

Tous les projets

Les projets IBM SPSS Modeler Advantage aussi bien que non IBM SPSS Modeler Advantage peuvent créer des noeuds terminaux distincts pour l'évaluation en temps réel et par lots. Lors de la planification de travaux via le client IBM SPSS Collaboration and Deployment Services Deployment Manager, assurez-vous de configurer l'étape de travail de manière que seul le noeud Export concerné par l'évaluation par lot ne soit exécuté.

Sélection de données et de sous-ensembles

Dans la liste de **Source de données à évaluer**, sélectionnez la source de données requise.

Sous-ensembles de données

Si vous ne souhaitez pas évaluer toute la source de données, vous pouvez sélectionner un sous-ensemble des données en utilisant une règle existante ou en créant une nouvelle règle qui servira uniquement pour ce modèle. Pour plus d'informations, voir la rubrique «Définition de règles de sélection», à la page 55.

Sélection des enregistrements à scorer

Pour les modèles qui renvoient des scores de propension, vous pouvez employer ces valeurs pour sélectionner les enregistrements à scorer. Par exemple, vous pouvez marquer 30 % des clients avec la propension à répondre la plus élevée et les utiliser pour générer une liste de diffusion. Les scores de propension indiquent la probabilité de la valeur "true" pour la cible et sont disponibles pour les modèles prédictifs avec des cibles indicateur, ainsi que pour les modèles de règle prédictifs (interactifs). (Dans le cas des modèles de règle prédictifs, la réponse sélectionnée est vraie (valeur "true") et toutes les autres réponses sont fausses. Pour d'autres types de modèle, ces options sont indisponibles.)

Il existe trois façons de sélectionner des enregistrements selon la propension :

- **Le meilleur pourcentage.** Par défaut, cette option sélectionne les premiers dix pour cent. Cette option permet de simuler les résultats d'évaluation en fonction du profit. Pour plus d'informations, voir la rubrique «Simulation du profit maximal», à la page 71.
- **Plus grand nombre d'enregistrements.** Par défaut, cette option sélectionne les 1000 premiers enregistrements, le cas échéant.
- **En indiquant les valeurs minimale et maximale de propension.** Par exemple, tous les enregistrements dont la propension à répondre est comprise entre 75 % et 100 %.

Sélection de la destination de l'évaluation (scoring)

Vous pouvez, au choix, envoyer les données de scoring dans une base de données, dans un fichier, sur un serveur analytique ou sur un serveur Cognos BI. Les types de fichier pris en charge sont notamment les suivants : texte (*.txt, *.csv), Excel, le format *.sav utilisé par IBM SPSS Statistics, SAS et les formats pris en charge par Data Collection. Pour plus d'informations, voir la rubrique «Destination Fichier», à la page 78.

Destination Base de données

Lorsque vous choisissez d'envoyer les données de sortie scorées à une base de données, vous êtes invité à entrer des détails supplémentaires.

1. Entrez le **nom de la base de données** ou recherchez-le à l'aide de la fonction Parcourir.
2. Indiquez le mode de stockage des scores. Vous pouvez créer une nouvelle table, auquel cas vous êtes invité à entrer son nom. Vous pouvez également ajouter les données à une table existante ou remplacer une source de données. Dans ce cas, vous êtes invité à sélectionner la table ou la source de données appropriée.
3. Modifiez la taille de chaîne de valeur par défaut, si nécessaire.

Lorsque vous utilisez une base de données existante pour l'ajout ou le remplacement de scores, vous pouvez mapper les champs d'évaluation aux champs de la destination existante. Pour plus d'informations, voir la rubrique «Mappage de champs d'évaluation», à la page 80.

Destination Fichier

Les scores peuvent être envoyés dans différents types de fichier. Pour chaque type, vous recherchez un fichier existant à l'aide de la fonction Parcourir ou vous créez un nouveau fichier pour y ajouter les données d'évaluation.

Lorsque vous utilisez un fichier existant pour l'ajout ou le remplacement de scores, vous pouvez mapper les champs d'évaluation aux champs de la destination existante. Pour plus d'informations, voir la rubrique «Mappage de champs d'évaluation», à la page 80.

Fichier Excel

Type. Sélectionnez le fichier de type Excel que vous souhaitez exporter.

Inclure des noms de champ. Indique si les noms de champ doivent être inclus dans la première ligne de la feuille de calcul.

Fichier plat

Si vous choisissez d'envoyer les données de sortie scorées vers un fichier à plat, vous pouvez entrer des données supplémentaires pour contrôler le contenu du fichier.

Mode écriture. Si vous sélectionnez **Remplacer**, les éventuelles données présentes dans le fichier indiqué seront écrasées. Si vous sélectionnez **Ajouter**, la sortie sera ajoutée à la fin du fichier et les données existantes conservées.

- **Inclure des noms de champ.** Si vous sélectionnez cette option, les noms des champs figurent sur la première ligne du fichier de sortie. Cette option est disponible uniquement avec le mode d'écriture **Remplacer**.

Nouvelle ligne après chaque enregistrement. Si vous sélectionnez cette option, chaque enregistrement est écrit sur une nouvelle ligne dans le fichier de sortie.

Séparateur de champ. Indique le caractère à insérer entre les valeurs des champs dans le fichier texte généré. Les options sont les suivantes : **Virgule**, **Tabulation**, **Espace** et **Autre**. Si vous sélectionnez **Autre**, entrez les caractères de séparation souhaités dans le champ de texte.

Guillemets. Indique le type de guillemet à utiliser pour les valeurs des champs symboliques. Les options disponibles sont les suivantes : **Aucun** (valeurs non accompagnées de guillemets), **Simple (')**, **Double (")** et **Autre**. Si vous sélectionnez **Autre**, entrez le type de guillemet souhaité dans le champ de texte.

Codage. Indique la méthode de codage de texte employée. Vous pouvez sélectionner la valeur par défaut du système ou UTF-8.

Symbole décimal. Indique le mode de représentation des décimales dans les données.

- **Point (.)**. Le point est utilisé comme séparateur décimal.
- **Virgule (,)**. La virgule est utilisée comme séparateur décimal.

Fichier de données IBM SPSS Statistics

Exporter les noms de champ. Indique une méthode de gestion des noms de variable et des libellés, applicable lors de l'exportation vers un fichier IBM SPSS Statistics *.sav*. Les noms incorrects sont automatiquement corrigés par le remplacement des caractères non valides.

- **Noms et libellés de variable.** Sélectionnez cette option pour exporter les noms de champ et les libellés de champ. Les noms sont exportés en tant que noms de variable IBM SPSS Statistics, alors que les libellés le sont en tant que libellés de variable IBM SPSS Statistics.
- **Noms comme libellés de variables.** Sélectionnez cette option pour utiliser les noms de champ comme libellés de variable dans IBM SPSS Statistics. Pour éviter de créer des noms IBM SPSS Statistics incorrects, sélectionnez plutôt **Noms et libellés de variable**.

Fichier de données SAS

Type. Sélectionnez le format de fichier SAS à créer. Vous pouvez sélectionner l'un des trois formats de fichier SAS disponibles : SAS pour Windows/OS2 (*.sd2), SAS pour UNIX (*.ssd) ou SAS version 7/8/9 (*.sas7bdat).

Exporter les noms de champ. Sélectionnez les options pour exporter les noms de champ et les libellés à utiliser avec SAS.

- **Noms et libellés de variable.** Sélectionnez cette option pour exporter les noms de champ et les libellés de champ. Les noms sont exportés en tant que noms de variable SAS, alors que les libellés le sont en tant que libellés de variable SAS.
- **Noms comme libellés de variables.** Sélectionnez cette option pour utiliser les noms de champ comme libellés de variable dans SAS. Pour éviter de créer des noms SAS incorrects, sélectionnez **noms** à la place.

Destination sur le serveur analytique

Lorsque vous choisissez d'envoyer les données de sortie scorées à un serveur analytique, vous êtes invité à entrer des détails supplémentaires.

1. Spécifiez la **source de données**.
2. Indiquez le mode de stockage des scores. Vous pouvez ajouter les données à une table existante ou remplacer une source de données. Dans ce cas, vous êtes invité à sélectionner la table ou la source de données appropriée.
3. Sélectionnez les champs de sortie à inclure avec la sortie d'évaluation.
4. Indiquez les enregistrements à évaluer. Dans la liste déroulante Evaluation, indiquez l'option d'évaluation choisie : **Tous les enregistrements**, **Top N%**, **Top n enregistrements**, **Propension Min/Max** (disponible uniquement pour les cibles *indicateur*) ou **Score minimum** (disponible uniquement pour les cibles *continues*).

Destination Serveur Cognos BI

Lorsque vous choisissez d'envoyer les données de sortie scorées à un serveur Cognos BI, vous êtes invité à entrer d'autres détails.

Remarque : vous ne pouvez exporter que des données relationnelles, pas de données OLAP.

Pour exporter les données vers Cognos BI, vous devez indiquer ce qui suit :

- **Connexion à Cognos.** Connexion au serveur Cognos BI.
- **Connexion ODBC.** Connexion au serveur de données Cognos utilisée par le serveur Cognos BI.

Les connexions doivent pointer vers la même base de données ; le nom d'utilisateur et le mot de passe requis pour que le serveur Cognos se connecte à ODBC doivent correspondre aux détails ODBC.

Vous exportez les données réelles vers le serveur de données et les métadonnées de package vers le serveur Cognos BI.

1. Sélectionnez la **Destination** Serveur Cognos BI et cliquez sur **Connecter** ; vous êtes invité à entrer les détails de la connexion serveur.

- a. Entrez l'**URL de serveur** du serveur IBM Cognos à partir duquel vous souhaitez importer ou exporter les données. Contactez votre administrateur système Cognos si vous ne savez pas quel URL choisir.
- b. Sélectionnez votre **Mode** de connexion. Pour vous connecter en tant qu'utilisateur spécifique (par exemple, en tant qu'administrateur), sélectionnez **Définir les informations d'identification** et entrez l'espace-noms IBM Cognos, le nom d'utilisateur et le mot de passe :
 - Entrez l'**Espace de noms** du fournisseur d'authentification de sécurité IBM Cognos pour vous connecter au serveur. Le fournisseur d'authentification permet de définir et de gérer les utilisateurs, les groupes et les rôles, ainsi que de contrôler le processus d'authentification.
 - Entrez le **Nom d'utilisateur** sous lequel vous souhaitez vous connecter au serveur.
 - Entrez le **Mot de passe** associé au nom d'utilisateur indiqué.

OU

Parcourez et sélectionnez les données d'identification IBM Cognos stockées dans IBM SPSS Collaboration and Deployment Services Repository.

- c. Vous pouvez également vous connecter sans les données d'identification utilisateur. Dans ce cas, vous n'indiquez pas l'espace de noms, le nom d'utilisateur ou le mot de passe, mais vous sélectionnez **Utiliser la connexion anonyme**.

Remarque : certaines connexions serveur n'autorisent pas les connexions anonymes.

2. Sélectionnez la **Source de données**.
3. Entrez le chemin et le nom du **Dossier** situé sur le serveur Cognos BI sur lequel le package d'exportation doit être créé. Vous pouvez également rechercher le dossier voulu à l'aide de la fonction Parcourir.
4. Entrez le **Nom du package** qui doit contenir les métadonnées exportées. Il doit s'agir d'un nouveau package ; vous ne pouvez pas exporter vers un package existant. Pour plus d'informations, voir la rubrique «Sélection des détails d'objet Cognos», à la page 43.
5. Pour sélectionner la destination ODBC, entrez le **Nom de la base de données** ou recherchez-le à l'aide de la fonction Parcourir.
6. Indiquez le mode de stockage des scores. Vous pouvez créer une nouvelle table, auquel cas vous êtes invité à entrer son nom. Vous pouvez également ajouter les données à une table existante ou remplacer une source de données. Dans ce cas, vous êtes invité à sélectionner la table ou la source de données appropriée.

Lorsque vous utilisez une base de données existante pour l'ajout ou le remplacement de scores, vous pouvez mapper les champs d'évaluation aux champs de la destination existante. Pour plus d'informations, voir la rubrique «Mappage de champs d'évaluation».

Mappage de champs d'évaluation

Lorsque vous utilisez une table de base de données ou un fichier existant pour l'ajout ou le remplacement de scores, vous pouvez mapper les champs d'évaluation aux champs de la table ou du fichier existant. La sortie et le champ mappé associé doivent avoir le même type de données.

- Lors de l'écriture de scores dans une table de base de données existante, tous les champs doivent être mappés à des champs existants pour que l'évaluation se poursuive, que les nouveaux scores soient substitués ou ajoutés aux anciens. Si nécessaire, cliquez sur **Mapper les champs** pour indiquer des sorties pour les champs qui subsistent.
 - Lors de l'écriture dans un fichier existant, le mappage n'est nécessaire que si **Ajouter** est sélectionné sous les options de fichier de sortie. Si vous remplacez un fichier existant, le mappage n'est pas nécessaire car le nouveau fichier remplace l'ancien.
1. Cliquez sur **Mapper les champs**. Une boîte de dialogue s'affiche. Le système effectue une meilleure supposition en mappant les champs pour vous. A tout moment, vous pouvez revenir à la meilleure supposition du système en cliquant sur **Réinitialiser**.
 2. Faites votre choix parmi les destinations de sortie disponibles et cliquez sur **Mapper** pour les mapper aux champs d'évaluation requis.

3. Lorsque vous avez terminé, cliquez sur **Enregistrer**.

Sélection de champs de sortie

Sur l'onglet Evaluation, sélectionnez **Indiquer les champs de sortie** pour sélectionner les champs à inclure avec les données de sortie scorées. Outre les scores réels, vous avez généralement besoin d'un champ au minimum pour identifier chaque enregistrement (par exemple, le champ qui contient l'identificateur de client). Tout ou partie des champs d'entrée utilisés dans la modélisation peuvent également être inclus.

Champs de sortie du modèle

Par convention, les noms des champs de sortie d'évaluation sont basés sur le nom de champ cible, auquel est ajouté un préfixe personnalisé. Par exemple si le champ cible est appelé *Response*, les champs de sortie porteront des noms tels que *\$XF-Response* et *\$XFRP-Response*. Les champs spécifiques dépendent du type de modèle et du niveau de mesure appliqué au champ cible sélectionné, comme détaillé ci-après. Pour plus d'informations, voir la rubrique «Niveaux de mesure», à la page 46.

Cibles indicateur. Les modèles automatisés avec cibles indicateur affichent les champs suivants :

- **\$XF-<target>**. Valeur prévue pour chaque enregistrement, rapportée en termes de valeurs "true" et "false" définies pour le champ.
- **\$XFRP-<target>**. Score de propension ; indique la probabilité de la valeur "true" pour chaque enregistrement. Les scores de propension diffèrent des cotes de confiance, qui sont typiquement rapportées en termes de prévision pour l'enregistrement concerné. Par exemple, une prévision fausse ("false") dotée d'un niveau de fiabilité élevé se traduit par une propension basse, car elle indique une forte probabilité de ne *pas* répondre.

Cibles catégorielles. Les modèles automatisés avec cibles catégorielles (nominales ou ordinales) affichent les champs suivants :

- **\$XS-<target>**. Valeur prévue pour chaque enregistrement.
- **\$XSC-<target>**. Valeur de confiance associée à la prévision.

Cibles continues. Les modèles automatisés avec cibles numériques continues affichent les champs suivants :

- **\$XR-<target>**. Valeur prévue pour chaque enregistrement.
- **\$XRE-<target>**. Erreur standard pour la prévision.

Modèles de règle prédictifs (interactifs). Les modèles de règle prédictifs (cibles indicateur ou catégorielles uniquement) affichent les champs suivants :

- **\$D-<target>**. Résultat prévu pour chaque enregistrement ; la valeur "true" est définie pour le champ utilisé pour indiquer la réponse souhaitée et NULL représente toute autre valeur.
- **\$DP-<target>**. Probabilité de résultat souhaité pour chaque enregistrement ; correspond à la probabilité affichée pour le segment dans la table de résultats lors de la génération du modèle. Pour plus d'informations, voir la rubrique «Génération d'un modèle de règle prédictif», à la page 61.
- **\$DI-<target>**. Numéro d'index utilisé pour identifier le segment dans lequel l'enregistrement se situe (s'affiche dans la colonne la plus à gauche dans les résultats de modélisation).
- **\$DRP-<target>**. Score de propension indiquant la probabilité de la réponse souhaitée. Disponible pour les cibles indicateur uniquement.

Evaluation de modèles

Une fois que vous avez indiqué les détails de l'évaluation, vous pouvez prévisualiser un échantillon des scores des modèles. Ceci est particulièrement utile avec un grand ensemble de données, car vous pouvez

rapidement vérifier la cohérence des résultats de l'évaluation par rapport aux résultats attendus, sans devoir attendre que tout l'ensemble des données soit évalué.

Remarque : si des données s'avèrent non valides ou incomplètes, un avertissement s'affiche lorsque vous tentez de les prévisualiser.

L'aperçu présente les données dans un tableau qui affiche les enregistrements par ligne et les variables par colonne.

Evaluation (scoring)

Pour déterminer le nombre d'enregistrements à évaluer, cliquez sur l'icône Actualiser. Vous obtenez alors le comptage du nombre d'enregistrements qui seront utilisés pour générer le modèle, en prenant en considération le nombre total d'enregistrements inclus dans la source de données ainsi que les paramètres facultatifs éventuellement indiqués.

Lorsque vous cliquez sur **Evaluer maintenant**, une barre de progression s'affiche dans la partie inférieure de l'onglet. Vous pouvez arrêter l'évaluation à tout moment pendant l'affichage de la barre de progression.

Remarques :

Si des données s'avèrent non valides ou incomplètes, un avertissement s'affiche lorsque vous tentez de les évaluer.

Lors de l'envoi du scoring dans une base de données, l'exécution est renvoyée par commande push à cette base de données si possible (cette opération est souvent appelée "SQL pushback"). Si le projet inclut une règle qui contient une référence à un modèle, l'exécution n'est pas renvoyée à la base de données.

Une fois que les données ont été évaluées avec succès, un message indique où les résultats sont stockés (dans la base de données ou dans le fichier que vous avez précisé) et la section Rapport récapitulatif d'évaluation propose un récapitulatif. Pour plus d'informations, voir la rubrique «Rapport récapitulatif d'évaluation». Si vous avez choisi d'envoyer les données de sortie scorées vers un fichier à plat et qu'un administrateur a activé la fonctionnalité correspondante, vous pouvez télécharger une copie locale de ce fichier.

Rapport récapitulatif d'évaluation

Après l'évaluation d'un projet, la section Rapport récapitulatif d'évaluation fournit des détails tels que l'emplacement cible des données évaluées, la taille du fichier (le cas échéant), le nom de la base de données et de la table (le cas échéant), le nom du modèle ou du projet, nombre d'enregistrements évalués, dans certains cas, un graphique et/ou un tableau récapitulatifs du modèle. La section Rapport récapitulatif d'évaluation est disponible dans IBM SPSS Modeler Advantage et sur l'onglet Déployer des applications qui incluent la fonction Evaluer maintenant.

Si vous choisissez d'envoyer les données de sortie scorées vers un fichier à plat, vous pouvez également télécharger une copie locale de ce fichier. Les administrateurs peuvent utiliser l'outil basé sur navigateur, IBM SPSS Collaboration and Deployment Services Deployment Manager, pour activer ou désactiver cette fonction, ainsi que pour augmenter ou réduire la taille de fichier maximale autorisée. Si le bouton **Télécharger le fichier** est inactif, cela signifie que la taille de fichier dépasse la limite maximale définie par l'administrateur ou que celui-ci a désactivé la fonction. La fonction est désactivée par défaut.

Chapitre 8. Administration des applications

Les administrateurs peuvent contrôler le mode d'interaction des utilisateurs métier avec les applications IBM Analytical Decision Management. Ce niveau de contrôle facilite les travaux des utilisateurs métier qui peuvent alors exécuter l'application de manière plus rapide et efficace dans la résolution des problèmes métier. Il permet également aux administrateurs de restreindre l'accès à certaines fonctionnalités.

Utilisez IBM SPSS Collaboration and Deployment Services Deployment Manager pour accorder des droits d'administration à des utilisateurs ou à des groupes spécifiques, selon vos besoins. Seuls les utilisateurs bénéficiant de l'action *Administration de Decision Management* ont accès aux fonctions décrites dans cette section.

1. Ouvrez IBM SPSS Collaboration and Deployment Services Deployment Manager.
2. Dans le menu Outils, sélectionnez **Server Administration (Administration de serveur)**.
3. Sur l'onglet d'administration de serveur, connectez-vous à votre serveur.
4. Sous **Utilisateurs et groupes**, sélectionnez **Référentiel d'utilisateurs local**.
5. Cliquez sur **Nouveau - Groupe** pour créer un groupe pour les administrateurs et un autre groupe pour les utilisateurs métier. Par exemple :
 - Decision Management Administrators (Administrateurs de Decision Management)
 - Decision Management Users (Utilisateurs de Decision Management)

Vous pouvez également créer de nouveaux utilisateurs à ce stade, si vous le souhaitez. Lors de la création de groupes, ajoutez les utilisateurs appropriés dans chaque groupe. Par exemple, les membres du groupe d'administrateurs vont bénéficier du droit d'administrer les applications IBM Analytical Decision Management dans l'organisation.

6. Accédez à Rôle et cliquez sur **Nouveau - Rôle** pour créer un rôle d'administration et un rôle utilisateur métier. Par exemple :
 - IBM Analytical Decision Management administrator (administrateur d'IBM Analytical Decision Management)
 - IBM Analytical Decision Management user (utilisateur d'IBM Analytical Decision Management)

Lors de la création de nouveaux rôles, affectez-leur au moins les actions suivantes :

Tableau 7. Actions associées aux rôles.

Rôle	Actions requises
Decision Management administrator (administrateur de Decision Management)	Accéder aux contenus et aux dossiers
	Administration de Decision Management
	Modification de travail
	Exécuter des rapports en mode dynamique
	Evaluer un modèle
	Soumettre un travail
	Afficher toutes les versions

Tableau 7. Actions associées aux rôles (suite).

Rôle	Actions requises
Decision Management user (utilisateur de Decision Management)	Accéder aux contenus et aux dossiers
	Modification de travail
	Exécuter des rapports en mode dynamique
	Evaluer un modèle
	Soumettre un travail
	Afficher la version la plus récente OU Afficher toutes les versions*

*Remarque : l'action *Afficher la version la plus récente* ou *Afficher toutes les versions* est obligatoire si les utilisateurs doivent accéder à des versions de projet qu'ils n'ont pas créées. Sinon, les projets s'ouvrent en mode lecture seule si l'utilisateur en cours n'a pas accès à la dernière version.

7. Sélectionnez le rôle "Decision Management administrator" et cliquez sur **Edit Users and Groups (Editer les utilisateurs et les groupes)**. Ajoutez le groupe "Decision Management Administrators" que vous venez de créer.
8. Sélectionnez le rôle "Decision Management user" et cliquez sur **Edit Users and Groups (Editer les utilisateurs et les groupes)**. Ajoutez le groupe "Decision Management Users" que vous venez de créer.

Options générales

Certaines options générales sont à la disposition des administrateurs sur tous les onglets de l'interface utilisateur d'IBM Analytical Decision Management. Ces options permettent aux administrateurs de masquer ou de verrouiller des onglets pour les utilisateurs métier.

- La partie supérieure de chaque onglet propose l'onglet **Masquer l'onglet <nom d'onglet> aux non administrateurs**, où <nom d'onglet> représente le nom de l'onglet. Cette option masque tout l'onglet aux utilisateurs métier. Par exemple, si vous ne voulez pas que les utilisateurs voient l'onglet Données pour une application, connectez-vous à l'application concernée, accédez à l'onglet Données et sélectionnez **Masquer l'onglet Données aux non administrateurs**.
- La partie supérieure de chaque onglet propose l'option **Verrouiller toutes les options de <nom d'onglet>**, où <nom d'onglet> représente le nom de l'onglet. Cette option verrouille tous les contrôles de l'onglet qui deviennent ainsi accessibles en lecture seule à l'utilisateur métier. Par exemple, si vous souhaitez que les utilisateurs métier puissent visualiser les paramètres de données en cours sans les changer, vous pouvez définir cette option sur l'onglet Données d'une application. Ce paramètre est également très utile si vous souhaitez préconfigurer certaines valeurs pour l'utilisateur métier afin de lui faciliter son travail.
- Les administrateurs peuvent également verrouiller et déverrouiller des paramètres individuels sur chaque onglet en cliquant sur l'icône de cadenas, si celle-ci est disponible.



Par exemple, dans les applications comportant un onglet Définir, les administrateurs peuvent verrouiller l'arborescence hiérarchique de dimension pour que les utilisateurs ne puissent pas la changer. Cliquez sur l'icône de cadenas pour verrouiller ou déverrouiller le contrôle associé. Les utilisateurs métier visualisent l'icône de cadenas verrouillée pour les contrôles verrouillés par l'administrateur, mais ils ne voient jamais l'icône de cadenas déverrouillée.

Les différents éléments verrouillés peuvent également s'afficher avec un arrière-plan grisé ou des couleurs amorties pour indiquer qu'ils ne sont pas disponibles à la sélection ; un message d'erreur s'affiche si un utilisateur tente de les sélectionner. Un message s'affiche dans la partie supérieure de l'onglet si celui-ci est entièrement verrouillé.

Contrôle des raccourcis d'application

La page de lancement d'application fournit aux utilisateurs métier une manière simple d'ouvrir des applications et d'ajouter des raccourcis personnalisés à certaines versions d'application marquées avec des libellés. Les administrateurs peuvent indiquer les raccourcis d'application à présenter sur la page de lancement pour tous les utilisateurs. Une fois connecté avec le droit d'accès administrateur :

- Si vous le souhaitez, cliquez sur l'icône de cadenas sur un raccourci d'application pour que les utilisateurs métier ne puissent pas le supprimer de leur page de lancement. Les utilisateurs métier peuvent toujours ajouter et supprimer d'autres raccourcis d'application. Les utilisateurs visualisent l'icône de cadenas verrouillée plutôt que l'icône Fermer pour les raccourcis verrouillés. Pour plus d'informations, voir la rubrique «Lancement d'application», à la page 27.

Masquage de la page d'accueil

Chaque application fournit un flux de travail étape par étape, représenté par des graphiques sur la page d'accueil. Cliquez sur un graphique pour passer à la section correspondante de l'interface utilisateur. Cependant, les utilisateurs métier n'ont pas toujours besoin d'avoir accès à chaque étape incluse dans le flux d'une application.

Les administrateurs peuvent masquer la page d'accueil aux utilisateurs métier. Par exemple, dans une application déterminée, les utilisateurs métier ne peuvent employer que l'onglet Rapports. Dans ce cas, un administrateur peut masquer aux utilisateurs métier la page d'accueil et tous les autres onglets. Ainsi, lorsque l'utilisateur métier se connecte, il accède directement à l'onglet Rapports et ne voit jamais les autres onglets.

1. Sur la page d'accueil, sélectionnez **Hide Application home page (Masquer la page d'accueil de l'application)**.
2. Dans la liste déroulante, indiquez l'onglet à afficher lorsque les utilisateurs ouvrent l'application pour la première fois.

Si vous masquez ultérieurement l'onglet sélectionné, l'utilisateur obtient l'affichage de l'onglet non masqué suivant dans l'application.

Verrouillage des options de données

Données

L'onglet Données permet de définir des ensembles de données pour l'analyse, la simulation, les tests, l'évaluation ou pour une autre utilisation opérationnelle. Ces sources de données sont définies par rapport à votre installation de serveur ; par conséquent, les chemins d'accès aux fichiers de données, aux sources ODBC, etc. sont tous relatifs au serveur.

En tant qu'administrateur, vous pouvez configurer tout ou partie des données avant que les utilisateurs métier ne commencent à exécuter les applications. Les administrateurs peuvent :

- Sélectionner **Verrouiller la source de données** lors de la création d'une nouvelle source de données pour la verrouiller une fois créée.
- Verrouiller et déverrouiller les différentes sources de données en cliquant sur l'icône de cadenas en regard d'une source de données dans la table, ou dans la boîte de dialogue Editeur de source de données.
- Définir et verrouiller le modèle de données utilisé dans l'application en sélectionnant un modèle de données dans la liste déroulante **Modèle de données du projet** et en cliquant sur l'icône de cadenas.

Les utilisateurs métier peuvent copier une source de données verrouillée de la section "Ensembles de données du projet" vers la section "Mes ensembles de données", mais elle reste verrouillée dans les deux ensembles. Pour plus d'informations, voir la rubrique Chapitre 4, «Gestion des sources de données», à la page 35.

- Lorsque le modèle de données du projet est une *vue de données*, utilisez le menu déroulant **Allow use of ODM rules based on hierarchical data (Autoriser l'utilisation de règles ODM basées sur des données hiérarchiques)** pour contrôler l'utilisation des règles IBM Operational Decision Management (ODM) qui présentent un modèle de données hiérarchique. L'utilisation de ces règles peut nuire aux performances. Les options suivantes sont disponibles dans la liste déroulante :
 - **Don't allow (Ne pas autoriser)** est recommandé pour de meilleures performances. Interdit dans tout le projet l'utilisation des règles ODM ayant une structure de données complexe. Une erreur s'affiche si un utilisateur sélectionne une règle complexe à un point quelconque dans l'application.
 - L'option d'**utilisation du seul onglet Modélisation** permet de faire référence aux règles complexes lors de la génération de modèle sur l'onglet Modélisation. Ceci permet d'ajouter une flexibilité supplémentaire, avec un impact sur les performances. Cette option s'applique seulement à IBM SPSS Modeler Advantage.

Verrouillage des options de destination d'évaluation

Dans les applications configurées pour utiliser la section Evaluer maintenant (onglet Déployer ou Evaluation), une section permet aux utilisateurs métier d'indiquer la destination où les données de score doivent être stockées (vers la base de données ou un fichier). Par exemple, si vous stockez les scores dans une base de données, vous indiquez le nom de base de données et la méthode de stockage des scores (comme créer une nouvelle table dans la base de données).

En tant qu'administrateur, vous pouvez verrouiller les options de destination d'évaluation. Par exemple, vous pouvez souhaiter uniquement ajouter les scores à une table existante d'une base de données déterminée, plutôt que d'accorder aux utilisateurs métier la possibilité de remplacer les données existantes, d'ajouter de nouvelles données à une table existante, de créer une nouvelle table ou de stocker des données dans un fichier.

- Pour verrouiller toutes les options de destination d'évaluation, cliquez sur l'icône de cadenas en regard de la liste déroulante Destination sur l'onglet Déployer. Les utilisateurs métier peuvent voir les options de destination, mais pas les éditer.

Téléchargement de métadonnées de projet à utiliser pour créer des règles externes

Les règles créées dans un système de gestion des règles métier tel qu'IBM Operational Decision Management peuvent être référencées et utilisées dans des applications IBM Analytical Decision Management si elles ont été développées pour prendre en charge le modèle de données employé dans le projet IBM Analytical Decision Management en cours. Pour ce faire, un administrateur IBM Analytical Decision Management doit télécharger les métadonnées appropriées. Il existe deux manières de télécharger les métadonnées ; la complexité du modèle de données et la complexité de la règle résultante potentielle diffèrent en conséquence :

Téléchargement des métadonnées de projet d'IBM Analytical Decision Management

Un administrateur peut cliquer sur l'icône **Télécharger les métadonnées** dans IBM Analytical Decision Management pour télécharger un fichier *.ZIP* contenant les métadonnées du projet en cours. Le fichier téléchargé est enregistré localement (sur le système utilisé par l'administrateur) pour déclencher le téléchargement.



Figure 14. Icône de métadonnées de téléchargement

Les métadonnées téléchargées de cette manière prennent en charge la création des règles basées sur le modèle de données du projet. Le modèle de données du projet sera non hiérarchique. Ainsi, toutes les règles créées sur la base de ces métadonnées peuvent seulement utiliser un modèle de données non hiérarchique. Les règles créées sur la base de ce modèle de données non hiérarchique peuvent être référencées dans un projet IBM Analytical Decision Management, quel que soit le type de la source de données du modèle de données du projet.

Les fichiers suivants sont inclus dans le fichier *zip* :

- **Définition de schéma XML (*.XSD).** Contient les définitions des types de champ du modèle de données pour le projet IBM Analytical Decision Management en cours. Ce fichier peut être importé dans un outil de développement externe tel qu'IBM Operational Decision Management Rule Designer afin de développer des règles utilisables avec le projet en cours.
- **Modèle Web Service Description Language (*.WSDL).** Inclus pour faciliter le développement de services Web qui utilisent le modèle de données. Le fichier *.WSDL* est fourni comme modèle utilisable pour générer des services Web chargés de la médiation des messages entre l'application IBM Analytical Decision Management et un service de règles externes. Les détails du service Web de médiation peuvent être manuellement fournis par l'intégrateur.

Téléchargement des métadonnées du modèle de données pour une vue de données analytique IBM SPSS Collaboration and Deployment Services

Dans IBM SPSS Collaboration and Deployment Services Deployment Manager, vous pouvez exporter des métadonnées pour une vue de données analytique en particulier.

Les métadonnées téléchargées de cette façon prennent en charge la création des règles basées sur le modèle de données logique d'une vue de données analytique. Le modèle de données logique d'une vue de données analytique peut être hiérarchique. Une source de vue de données IBM Analytical Decision Management prend en charge la génération d'une représentation XML des données hiérarchiques pour chaque enregistrement, qui peut alors être transmis à des règles externes pendant l'exécution. Les règles créées sur la base des métadonnées peuvent être complexes et utiliser le modèle de données hiérarchiques.

Dans IBM Analytical Decision Management, une règle externe basée sur un modèle de données hiérarchique peut seulement être référencée dans un projet dont le modèle de données utilise une source de vue de données. Pour assurer la compatibilité, la source de vue de données et la règle externe doivent être basées sur le modèle de données logique de la même vue de données analytique. Seule la source de vue de données peut générer les données hiérarchiques sur la base de l'exemple de vue de données analytique référencé. En d'autres termes, pour utiliser des règles externes, le modèle de données du projet IBM Analytical Decision Management doit correspondre à celui utilisé pour créer la règle, ce que permet la vue de données analytique.

Pour télécharger un fichier *.ZIP* avec des métadonnées pour le modèle de données logique de la vue de données analytique, procédez comme suit :

1. Dans IBM SPSS Collaboration et Deployment Services Deployment Manager, ouvrez une vue de données analytique à partir du référentiel et placez-vous dans l'onglet **Logique**.
2. Cliquez avec le bouton droit de la souris sur une table et exporter les métadonnées XOM.
3. Sélectionnez un emplacement pour enregistrer le fichier et cliquez sur **Enregistrer**.

Les fichiers suivants sont inclus dans le fichier *.ZIP* :

Définitions de schéma XML (*.XSD). Au moins deux définitions de schéma XML sont incluses. Une d'entre elle contient une définition qui peut être utile pour un type de réponse de règle. Les autres schémas contiennent les définitions des tables, y compris les détails de relation hiérarchique et les types de champ dans le modèle de données logique de la vue de données analytique.

Vous pouvez importer ces fichiers schémas dans un outil de développement externe comme IBM Operational Decision Management Rule Designer afin de développer des règles à utiliser avec un projet dont le modèle de données utilise une source de vue de données basée sur la même instance de vue de données analytique.

Pour obtenir des informations sur la création de règles externes à utiliser dans IBM Analytical Decision Management, voir *Application Designers Guide*.

Remarque : les administrateurs peuvent restreindre l'utilisation de ces règles externes complexes car elles peuvent dégrader les performances. Pour plus de détails, voir «Verrouillage des options de données», à la page 85.

Edition de la définition de schéma XML

Le fichier *XSD* téléchargé à partir d'IBM Analytical Decision Management contient deux types de données complexes : l'un représente le modèle de données du projet IBM Analytical Decision Management à utiliser pour le paramètre d'entrée des projets de règles IBM Operational Decision Management et l'autre, un type de retour de règles à employer pour le paramètre de sortie des projets de règles IBM Operational Decision Management.

Les métadonnées de vue de données analytiques IBM SPSS Collaboration and Deployment Services Deployment Manager contiennent au moins deux fichiers *.XSD*. Un de ces fichiers concerne le modèle de données logique d'une vue de données analytique, à utiliser pour le paramètre d'entrée des projets de règles IBM Operational Decision Management. Au moins un autre fichier contient un type complexe représentant un type de retour de règles, à utiliser pour les projets de règles IBM Operational Decision Management.

Les utilisateurs expérimentés peuvent choisir de modifier la définition de schéma *XSD* générée pour l'adapter aux règles spécifiques qu'ils souhaitent créer. Dans certains cas, avant de créer ses règles, l'utilisateur peut modifier les types liés à l'entrée et/ou à la réponse définis dans la définition de schéma *XSD* connexe. Par exemple, l'utilisateur expert doit parfois modifier le schéma lié au paramètre d'entrée pour que leurs règles écrites puissent prendre en charge le potentiel des champs d'entrée avec des valeurs NULL. Pour ce faire, l'utilisateur expert peut indiquer via la définition de schéma *.XSD* les éléments/champs de paramètre d'entrée obligatoires et facultatifs.

Edition du type de paramètre d'entrée

Le nombre d'occurrences possibles de chaque élément lié à un champ n'est pas explicitement défini dans le schéma *XSD* (les attributs `minOccurs` et `maxOccurs` ne sont pas définis ; leur valeur par défaut est 1). Si le schéma n'est pas édité avant d'être employé pour générer des règles IBM Operational Decision Management, tous les champs du type de paramètre d'entrée sont requis par les règles générées.

Modification du type de paramètre d'entrée associé au modèle de données du projet

Si le schéma définissant le type de paramètre d'entrée a été généré par téléchargement des métadonnées d'un projet IBM Analytical Decision Management, le modèle de données du paramètre d'entrée sera non hiérarchique. Dans ce cas, IBM Analytical Decision Management crée directement le contenu requis par le service de règle en cours d'exécution. Cela donne à l'utilisateur expert plus de liberté lors de la modification du type de paramètre d'entrée et de la conservation de la compatibilité. En d'autres termes, l'utilisateur expert peut retirer des champs/éléments inutiles pour la création de règles ou indiquer que des champs/éléments sont facultatifs.

Modification du type de paramètre d'entrée associé au modèle de données logique de la vue de données analytique

Si le schéma définissant le type de paramètre d'entrée a été généré par exportation du modèle de données logique d'une vue de données analytique à partir du client IBM SPSS Collaboration and Deployment Services Deployment Manager, le modèle de données du paramètre d'entrée peut être hiérarchique. S'il

est hiérarchique, IBM Analytical Decision Management ne créera *pas* directement le contenu requis par le service de règle en cours d'exécution. La partie principale du contenu sera générée par la source de vue de données analytique. Cela limite la possibilité, pour l'utilisateur expert, de modifier le type de paramètre d'entrée et de conserver la compatibilité. Par exemple, il peut uniquement indiquer que des champs/éléments sont facultatifs.

Informations générales

Avant d'appeler une règle créée dans un système de gestion des règles métier tel qu'IBM Operational Decision Management, IBM Analytical Decision Management demande au service de la règle si son modèle de données de paramètre d'entrée est hiérarchique ou non. Si le paramètre d'entrée est non hiérarchique, il vérifie également quelles sont entrées requises par le service de la règle.

Si le modèle de données du paramètre d'entrée est non hiérarchique et que l'enregistrement à évaluer contient une valeur NULL pour un des champs obligatoires du service de règle, IBM Analytical Decision Management n'appelle pas la règle et les champs de sortie/de réponse de celle-ci sont renseignés avec des valeurs par défaut. La valeur dépend du type du champ de sortie/réponse comme suit :

Tableau 8. Valeurs des types de champ de réponse.

Type de champ de réponse	Valeur
xsd:string	""
xsd:boolean	""
xsd:int	0
xsd:long	0
xsd:byte	0
xsd:double	0.0
xsd:float	0.0
xsd:date	0.0
xsd:time	0.0
xsd:datetime	0.0

Ceci peut avoir un impact sur les résultats des enregistrements d'évaluation via un service de règles. Par conséquent, soyez prudent lors de l'édition du paramètre d'entrée dans le schéma *XSD*.

Si le modèle de données du paramètre d'entrée est hiérarchique et que l'enregistrement à évaluer contient une valeur NULL pour les champs obligatoires du service de règle, IBM Analytical Decision Management appellera la règle mais le ou les champs avec une valeur NULL seront omis du contenu créé par le noeud source de vue de données. Si le champ omis était un champ obligatoire, cela entraînerait l'échec de l'appel de règle. Ainsi, il convient de prendre des précautions lors de la modification du paramètre d'entrée dans le schéma de *.XSD*.

La modification du schéma exige la connaissance des services de règles qui seront créés. Si un champ d'entrée déterminé n'est pas requis par la règle et que le modèle de données de paramètre d'entrée de cette même règle est **non hiérarchique**, l'élément lié au champ peut être **supprimé** du paramètre d'entrée dans le schéma ou être **marqué comme facultatif**. Si un champ particulier n'est pas requis par la règle et que le paramètre d'entrée de la règle est **hiérarchique**, l'élément lié au champ **ne peut pas être supprimé** mais peut être **marqué comme facultatif**. Effectivement, IBM Analytical Decision Management ne génère pas directement le contenu des règles basées sur un modèle de données hiérarchiques.

Si vous supprimez un champ de paramètre d'entrée, sa valeur est ignorée lorsque la règle est appelée, et celle-ci ne peut pas se rapporter au champ. Lorsque vous rendez un champ de paramètre d'entrée

facultatif, une valeur NULL est autorisée, et la règle doit être écrite défensivement pour le vérifier. Par exemple, si le paramètre d'entrée de la règle est non hiérarchique et si vous savez que le champ Age ne sera pas utilisé par la règle, vous pouvez supprimer l'élément connexe pour indiquer que ce champ n'est pas obligatoire. Dans l'exemple suivant de schéma XSD, si le champ Age n'est pas obligatoire, la section entière `<xsd:element name="Age">` peut être supprimée du schéma :

```
<xsd:complexType name="drugs">
<xsd:annotation>
  <xsd:appinfo>
    <dmname>drugs</dmname>
  </xsd:appinfo>
</xsd:annotation>
<xsd:sequence>
  <xsd:element name="Age">
    <xsd:annotation>
      <xsd:appinfo>
        <dmname>Age</dmname>
      </xsd:appinfo>
    </xsd:annotation>
    <xsd:simpleType>
      <xsd:restriction base="xsd:long">
        <xsd:minInclusive value="15"/>
        <xsd:maxInclusive value="74"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:element>
  <xsd:element name="Sex">
    <xsd:annotation>
      <xsd:appinfo>
        <dmname>Sex</dmname>
      </xsd:appinfo>
    </xsd:annotation>
    <xsd:simpleType>
      <xsd:restriction base="xsd:string">
        <xsd:enumeration value="F"/>
        <xsd:enumeration value="M"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:element>
  ...
</xsd:sequence>
</xsd:complexType>
```

Sinon, si le modèle de données du paramètre d'entrée de la règle est non hiérarchique ou hiérarchique, vous devrez sans doute rendre facultatif un élément de champ d'entrée, à condition qu'il ne soit pas requis par la règle. Pour cela, définissez la valeur d'attribut `minOccurs` de l'élément sur 0 :

```
<xsd:complexType name="drugs">
<xsd:annotation>
  <xsd:appinfo>
    <dmname>drugs</dmname>
  </xsd:appinfo>
</xsd:annotation>
<xsd:sequence>
  <xsd:element name="Age" minOccurs="0">
    <xsd:annotation>
      <xsd:appinfo>
        <dmname>Age</dmname>
      </xsd:appinfo>
    </xsd:annotation>
    <xsd:simpleType>
      <xsd:restriction base="xsd:long">
        <xsd:minInclusive value="15"/>
        <xsd:maxInclusive value="74"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:element>
  <xsd:element name="Sex">
    <xsd:annotation>
      <xsd:appinfo>
```

```

    <dmname>Sex</dmname>
  </xsd:appinfo>
</xsd:annotation>
<xsd:simpleType>
  <xsd:restriction base="xsd:string">
    <xsd:enumeration value="F"/>
    <xsd:enumeration value="M"/>
  </xsd:restriction>
</xsd:simpleType>
</xsd:element>
...
</xsd:sequence>
</xsd:complexType>

```

Exemple de modèle de données hiérarchiques :

```

<xsd:complexType name="Customer">
  <xsd:sequence>
    <xsd:element name="FirstName" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Surname" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Age" type="xsd:long"/>
    <xsd:element name="Sex" type="xsd:string"/>
    <xsd:element name="CustID" type="xsd:string"/>
    <xsd:element maxOccurs="unbounded" minOccurs="0" name="Orders" type="xom:Order"/>
  </xsd:sequence>
</xsd:complexType>
<xsd:complexType name="Order">
  <xsd:sequence>
    <xsd:element name="CustID" type="xsd:string"/>
    <xsd:element name="NoOfItems" type="xsd:long"/>
    <xsd:element name="TotalCost" type="xsd:long" minOccurs="0"/>
    <xsd:element name="Date" type="xsd:date"/>
  </xsd:sequence>
</xsd:complexType>

```

Les règles peuvent également être écrites défensivement pour manipuler des valeurs NULL. Dans ce cas, le champ peut être marqué comme facultatif (optional) dans le schéma. Voici un exemple de règle IBM Operational Decision Management nommée *currentOfferIsReplacePhone*, écrite pour traiter les valeurs NULL :

```

if
  the CURRENT OFFER of 'the customer' is not null
  and the CURRENT OFFER of 'the customer' is "Replace Phone"
then
  make it true that 'the response' is value ;
else
  make it false that 'the response' is value ;

```

Edition du type de paramètre de réponse

Il est possible de créer et d'utiliser des règles IBM Operational Decision Management pour renvoyer divers types à IBM Analytical Decision Management : *boolean* (booléen) utilisable dans les sélections, *numeric* (numérique) utilisable par exemple dans les règles d'agrégation et *string* (chaîne). Cependant, pour créer autre chose qu'une règle de réponse de données *string*, la définition de données de chaîne XML (.XSD) qui contient le type de réponse doit être modifiée manuellement.

En outre, un projet de règles peut contenir plusieurs règles, renvoyant chacune des types différents. Pour la prise en charge correspondante, il est nécessaire de gérer plusieurs versions du schéma XSD (par exemple, une version dans laquelle la variable data response (réponse de données) a été définie sur le type numérique, une autre version pour le type chaîne, une autre version pour le type booléen, etc.).

Le fichier XSD téléchargé à partir d'IBM Analytical Decision Management contient deux types de données complexes : l'un représente le modèle de données du projet IBM Analytical Decision Management à utiliser pour le paramètre d'entrée des projets de règles IBM Operational Decision Management et l'autre, un type de retour de règles à employer pour le paramètre de sortie des projets de règles IBM Operational Decision Management.

Le fichier *.XSD* lié à la réponse et téléchargé à partir d'IBM SPSS Collaboration and Deployment Services Deployment Manager contient la définition de type de réponse et ne contiendra que le type de données complexe.

Dans les deux cas ci-dessus, le type de retour (dont le nom contient toujours un suffixe "response") définit un élément *value*, lequel prend la valeur par défaut *string*. Si vous voulez créer un projet de règles avec un type différent de valeur de retour, vous devez éditer manuellement le schéma *XSD* téléchargé avant d'importer le projet de règles IBM Operational Decision Management, de sorte que l'élément *value* du type de retour soit défini selon les besoins. Voici quelques exemples :

```
<xsd:complexType name="claim_dataresponse">
  <xsd:annotation>
    <xsd:appinfo>
      <dmname>claim_dataresponse</dmname>
    </xsd:appinfo>
  </xsd:annotation>
  <xsd:sequence>
    <xsd:element name="value" type="xsd:string"/>
  </xsd:sequence>
</xsd:complexType>
```

```
<xsd:complexType name="claim_dataresponse">
  <xsd:annotation>
    <xsd:appinfo>
      <dmname>claim_dataresponse</dmname>
    </xsd:appinfo>
  </xsd:annotation>
  <xsd:sequence>
    <xsd:element name="value" type="xsd:boolean"/>
  </xsd:sequence>
</xsd:complexType>
```

```
<xsd:complexType name="claim_dataresponse">
  <xsd:annotation>
    <xsd:appinfo>
      <dmname>claim_dataresponse</dmname>
    </xsd:appinfo>
  </xsd:annotation>
  <xsd:sequence>
    <xsd:element name="value" type="xsd:long"/>
  </xsd:sequence>
</xsd:complexType>
```

Chapitre 9. Obtention de l'aide

Trois niveaux d'aide aux utilisateurs sont fournis :

- L'aide personnalisée accompagne chaque application préconfigurée fournie avec IBM Analytical Decision Management. Pour lancer l'aide, cliquez sur l'icône d'aide à partir de tout écran ou de toute boîte de dialogue :



- L'infobulle, qui s'affiche dans une petite fenêtre en incrustation, fournit des détails sur l'application ou le problème métier en cours. L'infobulle apporte une couche supplémentaire d'aide contextuelle personnalisée qui peut être facilement configurée pour un site ou une application spécifique. Elle apparaît dans toute l'interface utilisateur lorsque vous cliquez sur la petite icône d'infobulle :



- Infobulles standard, fournies pour les différents contrôles d'une page ou d'une boîte de dialogue. Pour afficher une infobulle, survolez le bouton, le lien ou autre contrôle.

Documentation disponible

La documentation d'IBM Analytical Decision Management est disponible dans deux formats. Lorsque vous accédez à l'aide dans l'application, l'IBM Knowledge Center est lancé, avec la rubrique d'aide pertinente affichée. Si vous éprouvez des difficultés à lancer l'aide, contactez votre administrateur au cas où vous ne bénéficieriez pas de droits d'accès à l'URL du Knowledge Center externe.

Le Knowledge Center est une collection centralisée en ligne de toute la documentation. Il est toujours disponible, même lorsque le produit n'est pas en cours d'exécution. Nous vous recommandons de marquer ce site d'un signet dans votre navigateur. Il contient également la documentation d'autres produits IBM, tels qu'IBM SPSS Modeler, IBM SPSS Collaboration and Deployment Services et Statistiques.

Si vous ne disposez pas de connexion Internet ou si le Knowledge Center est provisoirement inaccessible pour cause de maintenance, toute la documentation est également disponible au format (*.pdf). Les fichiers PDF sont disponibles avec les fichiers d'installation téléchargés, et sont également installés dans le répertoire d'installation d'IBM SPSS Collaboration and Deployment Services (par exemple, *C:\Program Files\IBM\SPSS\Deployment\8.0\Server\documentation\DecisionManagement*).

La documentation PDF est également disponible sur le Web à l'adresse <http://www-01.ibm.com/support/docview.wss?uid=swg27047202> et les notes sur l'édition sont disponibles à l'adresse <http://www-01.ibm.com/support/docview.wss?uid=swg27047203>.

Annexe A. Partage des flux

Les modèles et les projets créés dans IBM Analytical Decision Management sont enregistrés dans des fichiers de flux et peuvent être modifiés et utilisés dans IBM SPSS Modeler, si cette application est disponible. Par exemple, vous pouvez créer un flux de modèle ou de projet avec IBM SPSS Modeler Advantage ou une autre application IBM Analytical Decision Management, puis personnaliser le flux dans IBM SPSS Modeler avant d'effectuer un déploiement à l'aide d'IBM Analytical Decision Management. Il est également possible d'ouvrir dans IBM Analytical Decision Management des flux créés dans IBM SPSS Modeler, si des paramètres de déploiement ont été indiqués selon les instructions décrites dans cette annexe.

Les modèles qui respectent le format décrit dans cette annexe peuvent être visualisés, édités, évalués et actualisés à l'aide d'IBM SPSS Modeler Advantage. D'autres modèles peuvent également être évalués dans IBM SPSS Modeler Advantage, mais ils peuvent s'afficher de manière incorrecte et ils ne peuvent pas être actualisés.

Ouverture de flux dans IBM Analytical Decision Management

Lors de l'ouverture de flux dans IBM Analytical Decision Management ou IBM SPSS Modeler Advantage :

- Si le type de déploiement n'est pas précisé (Aucun), le flux peut être ouvert mais seul l'onglet Données est disponible.
- Lorsque le type de déploiement est **Evaluation seule**, l'onglet Evaluation peut également être disponible si la branche d'évaluation commence par un seul noeud de source de données dont le type est pris en charge par IBM SPSS Modeler Advantage ou IBM Analytical Decision Management. En d'autres termes, tous les types de source de données pris en charge par IBM SPSS Modeler le sont également par IBM SPSS Modeler Advantage et IBM Analytical Decision Management si le flux ne comporte qu'un seul noeud de source de données (aucun ensemble de fichiers fusionné).
- Lorsque le type de déploiement est **Actualisation de modèle**, l'onglet Modélisation peut également être disponible si la branche de modélisation commence par un seul noeud de source de données dont le type est pris en charge par IBM SPSS Modeler Advantage ou IBM Analytical Decision Management et que le noeud de modélisation sélectionné est de type suivant : un noeud de modélisation automatisé (pris en charge pour la modélisation prédictive), un noeud de liste de décision (pris en charge pour la modélisation de règle prédictive), un noeud de groupement (pris en charge pour la modélisation de classe) ou un noeud de règle d'association (pris en charge pour la modélisation d'association). En outre, un noeud de type doit être présent pour définir le champ cible ainsi que d'autres informations utilisées dans la modélisation.

Spécification des paramètres de déploiement dans IBM SPSS Modeler

Dans IBM SPSS Modeler, indiquez les paramètres de déploiement dans la boîte de dialogue Propriétés de flux, comme suit :

1. Dans les menus IBM SPSS Modeler, sélectionnez :
Outils > Propriétés de flux > Déploiement
2. Dans la liste de **Type de déploiement**, sélectionnez **Evaluation seule** ou **Actualisation de modèle**, selon le cas.
3. Sélectionnez un **Noeud d'évaluation**, un **Noeud de modélisation** et un **Noeud de modèle**, selon vos besoins.
4. Sélectionnez **Vérifier** pour confirmer que toutes les options obligatoires ont été indiquées ou sélectionnez **Stocker** pour stocker le flux dans IBM SPSS Collaboration and Deployment Services Repository.

Pour les flux créés dans IBM SPSS Modeler Advantage ou IBM Analytical Decision Management, les informations de déploiement sont automatiquement définies.

Astuces pour l'utilisation de flux IBM SPSS Modeler dans IBM Analytical Decision Management

- Il peut être plus facile de créer des flux avec IBM SPSS Modeler Advantage ou IBM Analytical Decision Management, puis de les modifier dans IBM SPSS Modeler, plutôt que de les générer entièrement à l'aide d'IBM SPSS Modeler. Les flux créés de cette manière sont structurés d'une manière qui leur permet d'être entièrement fonctionnels dans IBM Analytical Decision Management.
- Pour tous les flux créés dans IBM SPSS Modeler Advantage, des supernoeuds inclus à certains emplacements de la branche peuvent être personnalisés par un utilisateur expérimenté. Il est fortement recommandé d'effectuer les opérations de personnalisation ou d'ajout sur le flux dans ces supernoeuds pour s'assurer qu'elles sont conservées dans le flux et reconnues par IBM SPSS Modeler Advantage.
- Les noeuds de règles locales contenant une règle d'expression immédiatement en aval de chaque noeud source ont leurs champs inclus dans la boîte de dialogue Ajouter/Modifier des champs supplémentaires. Les règles locales peuvent être créés dans IBM Analytical Decision Management ou IBM SPSS Modeler Advantage, mais pas par IBM SPSS Modeler.
- Si un noeud de partition se trouve en amont du noeud de modélisation (et en aval du noeud de type), **Partitionner automatiquement les données pour activer le scoring du modèle dans la source de données de génération pour le scoring et le test** est activé dans IBM SPSS Modeler Advantage. L'état de la case à cocher est défini selon que le noeud de partition est activé ou non.
- Si un noeud de préparation automatique des données (ADP) se trouve en aval du noeud de type, alors **Nettoyer et préparer automatiquement les données pour la génération de modèle fiable** est activé. L'état de la case à cocher est défini selon que le noeud ADP est activé ou non.
- Si les noeuds Règle ou Sélection sont présents en amont du noeud de modélisation, ils sont utilisés pour remplir la section **Indiquer les sélections à utiliser**.
- Si des paramètres de flux ont été définis (onglet Paramètres dans la boîte de dialogue Propriétés de flux), un lien Paramètres s'affiche sur les onglets Modélisation et Evaluation dans IBM SPSS Modeler Advantage. L'utilisateur peut cliquer sur ce lien pour indiquer les valeurs ci-après.

Lancement d'IBM SPSS Modeler Advantage à partir d'IBM SPSS Modeler

A partir de l'écran d'accueil d'IBM SPSS Modeler, vous pouvez créer un modèle avec IBM SPSS Modeler Advantage.

Annexe B. Accessibilité

Les fonctions d'accessibilité facilitent aux utilisateurs à mobilité réduite ou malvoyants l'utilisation des produits de technologie de l'information. IBM s'efforce de concevoir des produits accessibles à tous les utilisateurs, quels que soient leur âge et leurs qualifications. Ce produit utilise les touches de navigation des navigateurs Web standard.

IBM Analytical Decision Management étant une application basée sur navigateur, les options d'accessibilité destinées aux utilisateurs atteints de handicaps, notamment aux personnes malvoyantes, sont définies à partir des paramètres de votre navigateur Web. Pour plus d'informations sur les actions de l'interface utilisateur (raccourcis clavier, lecteurs d'écran, etc.), reportez-vous à la documentation de votre navigateur Web.

Logiciels tiers

L'utilisation d'IBM Analytical Decision Management peut nécessiter l'emploi de logiciels tiers non couverts par le contrat de licence IBM. IBM décline toute responsabilité concernant les fonctions d'accessibilité de ces produits. Contactez le fournisseur pour obtenir des informations sur l'accessibilité de ses produits.

IBM et l'accessibilité

Pour plus d'informations sur l'engagement d'IBM en matière d'accessibilité, consultez le site IBM Human Ability and Accessibility Center.

Remarques

Il se peut qu'IBM propose ce matériel dans d'autres langues. Toutefois, il peut être nécessaire de posséder une copie du produit ou de la version du produit dans cette langue pour pouvoir y accéder.

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Pour plus de détails, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial IBM. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service IBM puisse être utilisé. Tout produit, programme ou service fonctionnellement équivalent peut être utilisé s'il n'enfreint aucun droit de propriété intellectuelle d'IBM. Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier lui-même les installations et applications réalisées avec des produits, logiciels ou services non expressément référencés par IBM.

IBM peut détenir des brevets ou des demandes de brevet couvrant les produits mentionnés dans le présent document. La remise de ce document ne vous donne aucun droit de licence sur ces brevets ou demandes de brevet. Si vous désirez recevoir des informations concernant l'acquisition de licences, veuillez en faire la demande par écrit à l'adresse suivante :

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
USA*

Les informations sur les licences concernant les produits utilisant un jeu de caractères double octet peuvent être obtenues en contactant IBM par écrit à l'adresse suivante :

*Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japon*

LE PRESENT DOCUMENT EST LIVRE "EN L'ETAT" SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Le présent document peut contenir des inexactitudes ou des coquilles. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, changer les produits et logiciels décrits dans ce document.

Les références à des sites Web non IBM sont fournies à titre d'information uniquement et n'impliquent en aucun cas une adhésion aux données qu'ils contiennent. Les éléments figurant sur ces sites Web ne font pas partie des éléments du présent produit IBM et l'utilisation de ces sites relève de votre seule responsabilité.

IBM pourra utiliser ou diffuser, de toute manière qu'elle jugera appropriée et sans aucune obligation à votre égard, tout ou partie des informations qui lui seront fournies.

Les licenciés souhaitant obtenir des informations permettant : (i) l'échange des données entre des logiciels créés de façon indépendante et d'autres logiciels (dont celui-ci), et (ii) l'utilisation mutuelle des données ainsi échangées, doivent adresser leur demande à :

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
USA*

Ces informations peuvent être soumises à des conditions particulières, prévoyant notamment le paiement d'une redevance.

Le logiciel sous licence décrit dans le présent document et tous les éléments sous disponibles s'y rapportant sont fournis par IBM conformément aux dispositions du Contrat sur les produits et services IBM, aux Conditions Internationales d'Utilisation de Logiciels IBM ou de tout autre accord équivalent.

Les données de performances et les exemples de clients ne sont présentés qu'à des fins d'illustration. Les performances réelles peuvent varier en fonction des configurations et des conditions d'exploitation spécifiques.

Les informations concernant des produits non IBM ont été obtenues auprès des fournisseurs de ces produits, par l'intermédiaire d'annonces publiques ou via d'autres sources disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances ni leur compatibilité. Aucune réclamation relative à des produits non IBM ne pourra être reçue par IBM. Toute question concernant les performances de produits non IBM doit être adressée aux fournisseurs de ces produits.

Les instructions relatives aux intentions d'IBM pour ses opérations à venir sont susceptibles d'être modifiées ou annulées sans préavis, et doivent être considérées uniquement comme un objectif.

Le présent document peut contenir des exemples de données et de rapports utilisés couramment dans l'environnement professionnel. Ces exemples mentionnent des noms fictifs de personnes, de sociétés, de marques ou de produits à des fins illustratives ou explicatives uniquement. Toute ressemblance avec des personnes ou des sociétés serait purement fortuite.

Marques

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines Corp., dans de nombreux pays. Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés. La liste actualisée de toutes les marques d'IBM est disponible sur la page Web "Copyright and trademark information" à l'adresse www.ibm.com/legal/copytrade.shtml.

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques commerciales ou déposées d'Adobe Systems Incorporated aux Etats-Unis et/ou dans d'autres pays.

Intel, le logo Intel, Intel Inside, le logo Intel Inside, Intel Centrino, le logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium, et Pentium sont des marques d'Intel Corporation ou de ses filiales aux Etats-Unis et dans certains autres pays.

Linux est une marque de Linus Torvalds aux Etats-Unis et/ou dans certains autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques de Microsoft Corporation aux Etats-Unis et/ou dans certains autres pays.

UNIX est une marque enregistrée de The Open Group aux Etats-Unis et/ou dans certains autres pays.

Java et toutes les marques et logos Java sont des marques commerciales ou déposées d'Oracle et/ou de ses affiliés.

Dispositions de la documentation du produit

Les droits d'utilisation relatifs à ces publications sont soumis aux dispositions suivantes.

Conditions d'utilisation

Les présentes dispositions viennent s'ajouter à toute autre condition d'utilisation applicable au site Web IBM.

Usage personnel

Vous pouvez reproduire ces publications pour un usage personnel et non commercial, sous réserve que toutes les mentions de propriété soient conservées. Vous ne pouvez distribuer ou publier tout ou partie de ces publications ou en faire des oeuvres dérivées sans le consentement exprès d'IBM.

Usage commercial

Vous pouvez reproduire, distribuer et afficher ces publications uniquement au sein de votre entreprise, sous réserve que toutes les mentions de propriété soient conservées. Vous ne pouvez reproduire, distribuer, afficher ou publier tout ou partie de ces publications en dehors de votre entreprise, ou en faire des oeuvres dérivées, sans le consentement exprès d'IBM.

Droits

Excepté les droits d'utilisation expressément accordés dans ce document, aucun autre droit, licence ou autorisation, implicite ou explicite, n'est accordé pour ces publications ou autres informations, données, logiciels ou droits de propriété intellectuelle contenus dans ces publications.

IBM se réserve le droit de retirer les autorisations accordées ici si, à sa discrétion, l'utilisation des publications s'avère préjudiciable à ses intérêts ou que, selon son appréciation, les instructions susmentionnées n'ont pas été respectées.

Vous ne pouvez télécharger, exporter ou réexporter ces informations qu'en total accord avec toutes les lois et règlements applicables dans votre pays, y compris les lois et règlements américains relatifs à l'exportation.

IBM N'OCTROIE AUCUNE GARANTIE SUR LE CONTENU DE CES PUBLICATIONS. LES PUBLICATIONS SONT LIVREES EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES PUBLICATIONS EN CAS DE CONTREFAÇON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE.

Index

A

- accessibilité 97
- administration
 - options de destination d'évaluation 86
 - options de données 85
 - options générales 84
 - page d'accueil 85
 - présentation 83
 - raccourcis d'application 85
- administration (présentation) 83
- Aide en ligne 93
- aider 93
- ajout de champs 48
- ajout de champs à l'aide de règles de segment 48
- ajout de champs supplémentaires 48
- ajout de formules 48
- ajouter de sorties de modèle 48
- ajouter/modifier des champs dérivés 48
- ajouter/modifier des champs supplémentaires 48
- annotations
 - ajout aux règles 57
- aperçu des données 45
- application, page d'accueil 28
- application, page de lancement 27, 85
 - ajout d'applications 27
 - personnalisation 27
- applications
 - lancement 13
 - test 72
- applications pré-générées 12
- appliquer des libellés aux versions 32
- artefacts de démonstration 11

B

- booléen 51

C

- champ cible
 - génération de modèles 15, 59
- champs
 - calcul 49
 - dans des expressions 50
 - mappage 47
- champs d'entrée
 - définition 45
 - mappage 47
- champs de sortie des modèles 81
- changer la langue 27
- changer la langue du navigateur 27
- classe manuelle 64
- contrôle
 - options générales 84
- contrôle des raccourcis d'application 85
- création de champs à l'aide de règles de segment 48

D

- date de simulation 29
- Dérivé, onglet 48
- dériver des champs 49
- description 33
- destination Base de données dans les modèles 77
- destination de l'évaluation dans les modèles 77
- destination Fichier dans les modèles 78
- destination Serveur Cognos BI dans les modèles 79
- destination sur le serveur analytique dans des modèles 79
- documentation disponible 93
- données d'échantillon 12
- données du serveur analytique 44
- données Hadoop 44
- données HDFS 44

E

- éditeur d'expression 49
- édition des valeurs de champ 47
- éléments inachevés
 - galerie 28
- enregistrement de fichiers 31, 32
- évaluation de modèles 17, 21, 69, 75, 81
 - destination Base de données 77
 - destination Fichier 78
 - destination Serveur Cognos BI 79
 - destination sur le serveur analytique 79
- mappage de champs 80
- modèles de groupement 22, 23
- modèles de liste interactive 20
- modèles de règle prédictif 20
- sélection de champs de sortie 81
- sélection de données 77
- sélection de la destination de l'évaluation (scoring) 77
- exemple de modélisation
 - d'association 23
- exemple de modélisation de groupement 22
- exemples d'applications 12
- exemples d'applications, artefacts 11
- exemples d'applications, données 12
- exemples d'applications, fichiers 11, 12
- exemples de fichiers 11
- exportation de règles 56
- expressions qui renvoient un booléen 51

F

- fichiers de données 12
- fichiers de données à base de texte 38
- fichiers de données Excel 38
- fichiers de données IBM SPSS Statistics 38

- flux

- flux
 - téléchargement 31
- fonctions
 - dans des expressions 50
- formats d'heure 29
- formats de date 29
- formats monétaires 29

G

- galerie
 - travail inachevé 28
 - travaux à exécution longue 28
- générateur de formules 48, 49, 51
- génération de modèles
 - champ cible 15
- graphique d'importance des prédicteurs 60
- graphique d'importance des variables 60
- graphique de fréquences 60
- graphiques dans les modèles 60
- graphiques de coïncidence 69
- graphiques de fréquences 69
- graphiques de fréquences des erreurs 69
- graphiques de gain 17, 69
- graphiques de lift 69
- graphiques de profits 17, 69
- graphiques de réponse 69
- graphiques de retour sur investissement 69

I

- IBM SPSS Collaboration and Deployment Services Repository
 - enregistrement dans 31, 32
- IBM SPSS Modeler Advantage
 - exemple 11
- icônes, Cognos BI IBM 44
- infobulle 93
- infobulles 93

L

- langue 27
- libellés de version 32

M

- mappage de champs 47, 80
- mappage de champs d'évaluation 80
- masquage de la page d'accueil 85
- mes sources de données 35
- modèle de données du projet 35
- modèles
 - association 64
 - champ cible 15, 59
 - définition des préférences 29
 - évaluation 17, 69

modèles (*suite*)
 évaluation (scoring) 21, 75, 81, 82
 génération 15, 59
 graphiques 60
 groupement 63
 classe manuelle 64
 limiter le temps de génération 29
 liste interactive 61
 mappage de champs d'évaluation 80
 paramètres facultatifs 65, 67
 prédictifs 59
 rapport récapitulatif d'évaluation 82
 règle prédictive 61
 résultats 60
 sélection d'enregistrements 77
 sélection de champs de sortie 81
 sélection de données 77
 sélection de la destination Base de données 77
 sélection de la destination de l'évaluation dans les modèles 77
 sélection de la destination Fichier 78
 sélection de la destination Serveur Cognos BI 79
 sélection de la destination sur le serveur analytique 79
 simulation du profit 71
 test 21, 72
 utilisation dans les règles 53
 modèles d'association 64, 67
 données tabulaires 65
 données tabulaires et données transactionnelles 65
 données transactionnelles 65
 modèles de groupement 63, 67
 modèles de liste interactive 61, 67
 modèles de règle prédictif 61, 67
 modèles prédictifs 59
 modification de champs dérivés 48
 modification des champs supplémentaires 48
 mots-clés 31, 33

N

niveaux de mesure 46
 Noeud source IBM Cognos BI
 icônes 44
 nombre maximum de membres pour les champs nominaux 33
 nouveautés 1, 2, 3
 nouvelles fonctions 1, 2, 3

O

obtention de l'aide 93
 documentation disponible 93
 opérateurs
 dans des expressions 50
 Operational Decision Management (ODM)
 règles 57
 options de destination d'évaluation 86
 options de données 85
 options générales 84
 ouverture de fichiers 31

P

page d'accueil 28, 85
 page de lancement 13, 27, 85
 ajout d'applications 27
 personnalisation 27
 paramètres
 dans des expressions 50
 partage de flux entre IBM Analytical Decision Management et IBM SPSS Modeler 95
 préférences 29
 préférences utilisateurs 29
 présentation
 administration 83
 présentation des données 45
 profit
 optimisation 11
 simulation 17
 profit, simulation dans les modèles 71
 propriétés
 description 33
 mots-clés 33
 nombre maximum de membres pour les champs nominaux 33
 projet 33
 référentiel 33
 publipostage, exemple 11

R

raccourcis
 application 27
 raccourcis d'application 27, 85
 rapport récapitulatif d'évaluation 82
 règles 53
 annotations 57
 créer 53
 exportation 56
 externes 57
 IBM Operational Decision Management 57
 logique de sélection 55
 partagées 56, 57
 réutilisation 56, 57
 segment 53
 sélection 55
 utilisation de modèles 53
 règles d'exclusion 55
 règles d'inclusion 55
 règles de segment
 utilisation de modèles 53
 règles de sélection 55
 règles externes
 utilisation dans des applications 57
 règles IBM Operational Decision Management (ODM)
 utilisation dans des applications 57
 règles métier 53
 règles partagées 56

S

sélection d'enregistrements dans les modèles 77
 sélection de données dans les modèles 77

simulation
 date par défaut 29
 simulation du profit dans les modèles 71
 source base de données 41
 sources Cognos BI 42
 sources de données
 à base de texte 38
 ajout de champs 48
 ajout de formules 48
 ajouter de sorties de modèle 48
 aperçu 45
 base de données 41
 calcul des champs 48
 champs d'entrée 45
 création de champs à l'aide de règles de segment 48
 définition 35, 37
 Dérivé, onglet 48
 édition des valeurs de champ 47
 Excel 38
 générateur de formules 49, 51
 Hadoop 44
 HDFS 44
 IBM Cognos BI 42
 IBM SPSS Statistics 38
 mappage de champs 47
 niveaux de mesure 46
 présentation des données 45
 serveur analytique 44
 sources Fichier 38
 vue de données 42
 sources de données du projet 35
 sources IBM Cognos BI 42

T

téléchargement de fichiers de flux 31
 téléchargement de flux 31
 test d'un modèle 72
 test d'une application 72
 test des modèles 21
 Types de données 46

U

URL
 application 27
 URL de l'application 27

V

valeurs globales
 dans des expressions 50
 vérification des expressions 51
 verrouillage des options de destination d'évaluation 86
 verrouillage des options de données 85
 verrouillage des projets 33
 vue de données 42

