



Points clés :

La norme CRISP-DM (CRoss-Industry Standard Process for Data Mining) de data mining sert de ligne directrice pour vous aider à évaluer les spécifications requises pour l'outil de data mining de votre entreprise.

Evaluation d'un outil de data mining

Commencez par définir le cadre global de votre évaluation

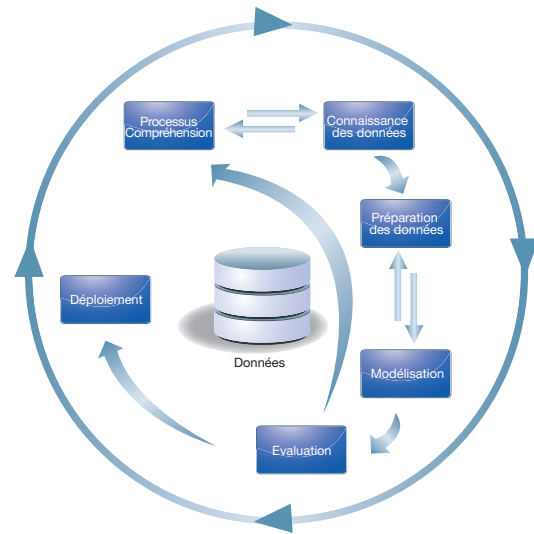
Le data mining vous aide à prendre de meilleures décisions qui débouchent sur des résultats concrets et riches de sens, se traduisant par exemple par une augmentation du chiffre d'affaires et des processus plus efficaces. Le data mining est extrêmement prometteur, mais la médiatisation dont il fait l'objet pourrait faire croire qu'il est possible d'obtenir des résultats incroyables moyennant seulement des efforts très modestes. Choisir un outil de data mining en se fondant sur cet espoir peut se solder par un retour sur investissements décevant.

Pour obtenir les résultats espérés d'un projet de data mining, faites une estimation de votre situation métier et évaluez les performances des outils tout au long du processus de data mining. Pour faciliter votre évaluation, la liste de contrôle suivante a été compilée à l'aide de CRISP-DM (Cross-Industry Standard Process for Data Mining).

CRISP-DM organise le processus de data mining en six phases, comme le montre le diagramme de la page suivante. La séquence des phases n'est pas rigide : il est toujours nécessaire d'effectuer des allers et retours entre les différentes phases. Gardez à l'esprit cet aspect interactif du data mining lors de votre évaluation de l'outil. L'outil que vous choisissez doit être assez souple pour que vous puissiez apporter des modifications à n'importe quelle phase au fur et à mesure que votre projet progresse.

Les critères d'évaluation qui suivent sont organisés selon ces six phases, et comportent à la fin des points supplémentaires à prendre en compte.





La Phase 1 vous aide à définir vos objectifs de projet et à élaborer un plan préliminaire de data mining.

CRISP-DM – Phase 1 : Compréhension de l'activité métier

La phase initiale se focalise sur la compréhension des objectifs et des exigences du projet du point de vue métier. Elle consiste ensuite à exprimer ces connaissances sous la forme d'une définition de problème de data mining et d'un plan préliminaire. La liste de contrôle permet de vérifier si l'outil a déjà été utilisé dans un large éventail de problèmes métier, et examine la façon dont il relie la compréhension de l'activité avec les aspects techniques du processus de data mining.

- L'outil a-t-il déjà permis de résoudre une gamme étendue de problèmes métier, notamment les problèmes auxquels nous sommes confrontés ?
 - L'outil s'est-il révélé utile pour résoudre des problèmes dans notre secteur d'activité ou avec notre application ? L'outil a-t-il déjà donné des résultats concluants avec les applications que nous voulons utiliser ?
- Comment l'outil fait-il le lien entre la compréhension métier et l'aspect technique du data mining ?
 - Les étapes d'utilisation de l'outil peuvent-elles être mises clairement en relation avec les besoins métier du data mining ? Les concepts du data mining sont-ils clairement définis pour l'utilisateur ? Comment l'outil s'intègre-t-il avec la gestion de projet ou les autres outils de planification ? Est-il nécessaire de créer des applications pour l'outil pour combler cette éventuelle lacune ?

La phase 2, permet de comparer la façon dont les outils de data mining que vous envisagez peuvent s'intégrer à votre environnement technologique actuel, en vue de mettre en œuvre des fonctions de collecte des données et de visualisation.

A la phase 3, vous évaluez la façon dont les différents outils de data mining préparent les données en vue d'une analyse optimale.

CRISP-DM – Phase 2 : Compréhension des données

La phase de compréhension des données commence avec la collecte des données et se poursuit par des activités permettant de se familiariser avec les données. Les normes d'accès aux données et les techniques de visualisation sont importantes au cours de cette phase.

- Comment l'outil préserve-t-il mon investissement existant dans l'infrastructure informatique ?
 - L'outil de data mining fonctionne-t-il bien avec nos entrepôts de données existants ? Est-il basé sur des normes ouvertes ou faudra-t-il faire migrer les données dans un format propriétaire ?
- Dans quelle mesure permet-il une exploration et une visualisation interactive des données ?
 - Fournit-il des techniques de visualisation permettant de déceler des modèles dans les données ? La visualisation peut-elle être effectuée en mode interactif, en apportant des modifications dans les graphiques eux-mêmes et en créant de nouveaux graphiques basés sur des dimensions différentes des données ?

CRISP-DM – Phase 3 : Préparation des données

La phase de préparation des données englobe toutes les activités de construction du jeu de données final à partir des données brutes de départ. Les éléments de la liste de contrôle incluent l'efficacité et la facilité de la préparation des données.

- Comment l'outil prépare-t-il les données ?
 - Le processus de data mining interactif pris dans sa totalité, y compris les activités de préparation des données brutes jusqu'à ce qu'elles soient prêtes pour la création de modèle, est-il adapté à un data mining efficace ? L'outil présente-t-il les étapes de la préparation des données d'une façon facile à suivre ?
- L'outil extrait-il automatiquement les données pour la préparation ?
 - L'outil extrait-il les données automatiquement ou l'écriture des requêtes SQL pour les jointures, les agrégations, les tris et les autres opérations de préparation des données doit-elle être effectuée manuellement ?

Une modélisation efficace est une composante essentielle du data mining. La Phase 4 vous aide à examiner les outils de data mining qui fonctionnent le mieux avec les différentes techniques de modélisation.

CRISP-DM – Phase 4 : Modélisation

Dans cette phase, différentes techniques de modélisation sont sélectionnées et appliquées et les paramètres sont calibrés de façon à obtenir des valeurs optimales. Souvent, les analystes reviennent à la phase de préparation des données pour satisfaire les exigences des différents types de modèles. Comme les différents modèles permettent de résoudre le même type de problème de data mining, l'un des principaux éléments de la liste de contrôle concerne l'aptitude de l'outil à appliquer et à comparer les différentes techniques.

- Améliore-t-il la productivité des analyses ?
 - L'outil permet-il aux analystes de développer rapidement des modèles efficaces ? Avec quelle facilité les utilisateurs peuvent-ils essayer différents modèles afin de déterminer la meilleure solution ? Est-il facile de préparer les données pour répondre aux besoins d'un modèle spécifique ?
- Offre-t-il un large choix de techniques ?
 - L'outil propose-t-il des techniques ou des algorithmes de visualisation, de classification, de groupement, d'association et de régression ?
- Permet-il de combiner différentes techniques ?
 - Différentes techniques peuvent-elles être facilement combinées pour obtenir de meilleurs résultats ? Les résultats des algorithmes peuvent-ils être incorporés au jeu de données à des fins de post-traitement et d'analyse ?
- Protège-t-il mes investissements actuels dans des technologies comme les algorithmes et les autres outils ?
 - L'outil de data mining peut-il fonctionner avec vos algorithmes existants ? Fonctionne-t-il bien avec vos autres outils d'analyse des données ?

A la phase 5, qui est la phase d'évaluation, vous examinez les différents modèles de données pour vous assurer que tous les problèmes métier ont été résolus.

CRISP-DM – Phase 5 : Evaluation

La phase d'évaluation est une évaluation minutieuse du ou des modèles avant le déploiement. Un objectif majeur consiste à déterminer si un problème métier important n'a pas été pris assez en considération. Parmi les éléments de la liste de contrôle utilisée dans la phase d'évaluation, figurent la bonne intégration des informations fournies par les utilisateurs, et la clarté des résultats pour le public visé.

- Donne-t-il en permanence de bons résultats ?
 - Crée-t-il des solutions caractérisées par un niveau de qualité élevé ou ne fournit-il de bons résultats que dans certains cas ou avec certains types de données ? Vos résultats reflètent-ils avec exactitude tous les problèmes métier et par conséquent, donnent-ils de bons résultats sur les données de test ?
- L'outil donne-t-il des résultats faciles à comprendre ?
 - Vos résultats sont-ils faciles à comprendre par les utilisateurs ? Sinon, comment devez-vous procéder pour rendre les résultats plus convaincants ? L'outil a-t-il incité les experts métier à communiquer des informations pendant tout le processus de data mining ?
- Toute la gamme de la visualisation peut-elle être appliquée afin de valider les résultats d'un modèle ?
 - Les prévisions, les scores et les autres résultats du modèle sont-ils faciles à analyser et permettent-ils de valider l'efficacité du modèle ?

CRISP-DM – Phase 6 : Déploiement

Le processus de data mining est utilisé au cours de la phase de déploiement, qu'il s'agisse simplement d'obtenir un éclairage sur un problème métier ou de l'implémenter dans une application pour fournir des connaissances actualisées aux destinataires de l'information. Souvent, le déploiement nécessite une grande quantité de services. Le principal élément de la liste de contrôle concerne donc la capacité de l'outil à faciliter cette tâche.

- Comment puis-je déployer mes solutions de data mining, dans l'immédiat et à l'avenir ?
 - Comment les solutions de data mining peuvent-elles être intégrées aux applications opérationnelles ? L'intégration peut-elle être réalisée à moindre coût, ou nécessite-t-elle des investissements importants en termes d'argent et de développement ? Mes solutions peuvent-elles être mises à jour facilement ? Si mes solutions ne peuvent pas être facilement mises à jour, quelle charge de travail et quels coûts cela implique-t-il ?

La Phase 6, de la liste de contrôle CRISP-DM vous aide à évaluer la facilité et l'efficacité du déploiement de solutions de data mining.

Autres facteurs à prendre en compte : le coût d'acquisition

Outre les phases de data mining du modèle de processus CRISP, il est nécessaire d'effectuer une analyse du retour sur investissements.

- Quel est le coût d'acquisition ?
 - Quantifiez les coûts de propriété pendant toute la durée de vie du produit et du service, sans omettre les systèmes complémentaires requis. Quantifiez le retour anticipé. A quel moment espérez-vous un retour sur investissements positif et ce résultat satisfait-il les objectifs métier ?
- Quelle est la durée d'implémentation ?
 - Combien de temps vous faudra-t-il pour implémenter votre outil de data mining ? Nécessite-t-il d'autres outils ou matériels ? Quel est l'effort de formation, de conseil et de personnalisation requis pour obtenir des résultats ?
- Les compétences des utilisateurs (actuelles et futures) correspondent-elles à celles requises pour utiliser l'outil ?
 - Quelles sont les compétences requises pour obtenir des résultats utiles avec l'outil de data mining ? Est-ce un outil conçu pour des utilisateurs très bien formés sur le plan technique, pour des débutants en data mining, ou peut-il être utilisé par des groupes et des personnes ayant un niveau de compétences intermédiaire ? Quels sont les coûts de formation à prévoir pour former tous les utilisateurs ? N'oubliez pas non plus de prendre en compte les compétences des futurs utilisateurs potentiels.
- L'outil peut-il être personnalisé pour nos utilisateurs ?
 - Comment l'outil peut-il être personnalisé en fonction d'usages différents ? Est-il possible d'enregistrer des processus courants en vue de les réutiliser ? L'outil peut-il automatiser des tâches ? Existe-t-il des services permettant de personnaliser les interfaces ou susceptibles de fournir un autre type d'assistance ?

Lorsque vous déterminez l'efficacité d'un outil de data mining, veillez à prendre en compte le coût total de possession, le retour sur investissements et les opportunités ultérieures de partenariat avec un fournisseur.

Autres facteurs à prendre en compte : le fournisseur

Pour conclure, vous devez réfléchir à l'importance du rôle du fournisseur dans votre évaluation d'outil de data mining. Par bien des façons, l'achat d'une solution de data mining est un investissement pour l'avenir. Il est donc conseillé de choisir un partenaire sûr pour vous accompagner dans votre projet.

- Le fournisseur propose-t-il d'autres outils et services susceptibles de vous aider à résoudre des problèmes similaires ?
 - Le fournisseur propose-t-il d'autres outils de data mining ou d'analyse des données ? Propose-t-il des prestations de conseil, de formation, de support technique et d'autres services ? Ces services sont-ils accessibles dans tous les pays ?
- Le fournisseur est-il une source fiable, et est-il capable d'actualiser ses logiciels et ses services afin de répondre à mes besoins ?
 - Le fournisseur est-il un leader en solution de data mining ? Possède-t-il les ressources requises pour continuer à fournir un niveau de service élevé à l'avenir ?

À propos d'IBM Business Analytics

Les logiciels IBM Business Analytics aident les entreprises à mesurer, comprendre et anticiper leur performance financière et opérationnelle en fournissant des informations exactes, cohérentes et complètes. Ce portefeuille complet de solutions décisionnelles, d'analyse prédictive, de gestion des performances financières et de stratégies et d'applications d'analyse fournit une vision claire, immédiate et concrète des performances actuelles et permet de prévoir les résultats futurs. Alliant de riches solutions sectorielles, des services professionnels et les meilleures pratiques, la solution permet aux entreprises de toutes tailles d'optimiser leur productivité, d'automatiser en toute confiance leurs décisions et de générer de meilleurs résultats.

Faisant partie de ce portefeuille, la solution logicielle IBM SPSS Predictive Analytics permet aux entreprises de prévoir les événements et d'agir proactivement en fonction des données obtenues pour obtenir de meilleurs résultats métier. Les clients des secteurs privé, public et universitaire du monde entier font appel à la technologie IBM SPSS, qui les dote d'un atout concurrentiel pour attirer, fidéliser et développer leur clientèle, tout en réduisant les fraudes et en atténuant les risques. En intégrant les logiciels IBM SPSS à leurs opérations quotidiennes, les entreprises deviennent proactives, capables de diriger et d'automatiser leurs décisions pour atteindre leurs objectifs métier et se doter d'un atout concurrentiel. Pour plus d'informations ou pour contacter un interlocuteur IBM, visitez le site ibm.com/spss/fr



Compagnie IBM France

17 Avenue de l'Europe
92 275 Bois-Colombes Cedex

La page d'accueil d'IBM est accessible à l'adresse

ibm.com

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines aux Etats-Unis et/ou dans certains autres pays. Si ces marques ou d'autres termes relatifs aux marques IBM apparaissent lors de leur première occurrence dans ce document accompagnés d'un symbole de marque (® ou ™), ces symboles indiquent qu'il s'agit de marques déposées aux Etats-Unis ou reconnues par la législation générale comme étant la propriété d'IBM au moment de la publication de ce document. Ces marques peuvent également exister et éventuellement avoir été enregistrées dans d'autres pays. La liste actualisée de toutes les marques d'IBM est disponible sur la page Web « Copyright and trademark information » à l'adresse ibm.com/legal/copytrade.shtml

SPSS est une marque de SPSS Inc., une société du groupe IBM, dans de nombreux pays.

© Copyright IBM Corporation 2012



Merci de recycler ce document

Business Analytics software