

L'utilisation du data mining pour détecter la fraude à l'assurance



Améliorer la précision et minimiser les pertes

Points clés :

- IBM SPSS Modeler associe de puissantes techniques analytiques aux efforts existants de prévention et de détection des fraudes
 - Créez des modèles basés sur des demandes d'indemnisation précédemment vérifiées et utilisez-les pour identifier de futures demandes potentiellement frauduleuses
 - Veillez à ce que vos experts se concentrent sur les demandes qui sont susceptibles d'être frauduleuses et revues à la baisse
 - Déployez les résultats vers les personnes pouvant utiliser ces informations pour éradiquer la fraude et récupérer de l'argent
-

Introduction

Chaque organisation qui échange de l'argent avec des clients, des prestataires de services ou des fournisseurs risque d'être victime de fraudes et d'abus.

Partout dans le monde, les compagnies d'assurance perdent de plus en plus d'argent chaque année à cause de demandes d'indemnisation frauduleuses. Il faut qu'elles récupèrent cet argent afin de pouvoir continuer à fournir des services de qualité à leurs clients.

Les outils de data mining d'IBM SPSS sont basés sur les normes du secteur et permettent aux agences d'associer le data mining IBM SPSS aux efforts existants de prévention et de détection des fraudes afin d'améliorer la précision, de réduire la main d'œuvre et de minimiser les pertes. L'effort conjugué d'IBM et de SPSS vous offre une flexibilité optimale et s'adapte aux types de données que vous explorez et à la façon dont vous déployez les résultats.

Pour s'assurer que les experts ciblent les demandes les plus susceptibles d'être revues à la baisse, de nombreuses compagnies d'assurance ont intégré le data mining dans leurs processus d'investigation et de vérification. Le data mining associe de puissantes techniques analytiques à vos connaissances commerciales pour transformer les données que vous avez déjà acquises en informations utiles pour identifier les cas probables de fraude et d'abus.

SPSS a fait partie des pionniers dans le domaine de l'analyse des données ; elle a été la première sur ce créneau et reste l'une des applications logicielles les plus populaires et les plus largement utilisées. En tant que nouveau membre de l'organisation IBM, SPSS met ses outils d'analyse de pointe à la portée d'un plus grand nombre de clients à travers le monde.

Les offres d'IBM SPSS incluent des produits de référence dans le domaine de la collecte de données, des statistiques et du data mining, avec une plateforme unifiée prenant en charge la gestion et le déploiement sécurisés des ressources d'analyse.



Récapitulatif :

Chaque année, les compagnies d'assurance perdent des millions de dollars à cause de demandes d'indemnisation frauduleuses, en grande partie parce qu'elles ne sont pas à même de distinguer facilement les demandes légitimes de celles qui peuvent être frauduleuses. Pour s'assurer que les experts ciblent les demandes les plus susceptibles d'être revues à la baisse, de nombreuses compagnies d'assurance ont intégré le data mining d'IBM SPSS à leurs processus d'investigation et de vérification. Ce document explique comment les techniques de data mining peuvent vous permettre d'améliorer la précision et d'économiser du temps, de l'argent et des ressources.

Découvrez comment éviter les pertes dues à la fraude

Comment votre organisation détermine-t-elle quelles demandes d'indemnisation sont légitimes parmi les milliers, voire les millions, qu'elle reçoit ? Peut-être que vos experts ont tendance à cibler les déclarations de sinistre ou les demandes de paiement donnant lieu à des rectifications négligeables, tout en passant à côté des cas qui permettraient de récupérer des sommes importantes. Et si vous pouviez :

- Découvrir de petits sous-ensembles de demandes ayant un fort pourcentage de fraude récupérable ?
- Isoler les facteurs qui indiquent qu'une déclaration de sinistre ou une demande de paiement risque fort d'être frauduleuse ?
- Développer des règles et les utiliser pour signaler uniquement les déclarations de sinistre ou les demandes d'indemnisation les plus susceptibles d'être frauduleuses ?
- Veiller à ce que vos experts puissent examiner les déclarations de sinistre ou les demandes d'indemnisation qui sont les plus susceptibles non seulement d'être frauduleuses mais aussi d'être revues à la baisse ?

Si votre compagnie d'assurance pouvait atteindre ces objectifs, vous pourriez faire un meilleur usage de vos ressources et prévenir et réduire plus efficacement les activités frauduleuses. Votre service pourrait alors réduire les importantes pertes d'argent encourues chaque année à cause des demandes d'indemnisation frauduleuses.

Exploitez les données existantes

Vos demandes d'indemnisation précédemment vérifiées sont la clé pour récupérer de l'argent à l'avenir. En créant des modèles à partir d'informations historiques, vous pouvez repérer avec exactitude les demandes frauduleuses parmi les millions que vous recevez chaque année. Ces modèles de data mining réduisent le coût de la fraude et des abus tout en faisant gagner du temps à vos experts. Le data mining permet à divers prestataires d'assurance de prédire quelles demandes sont frauduleuses afin de pouvoir cibler efficacement leurs ressources et récupérer d'importantes sommes d'argent.

Le scénario suivant montre comment un prestataire d'assurance (dans cet exemple un prestataire d'assurance maladie) a utilisé le data mining d'IBM SPSS pour créer des modèles basés sur des demandes d'indemnisation déjà vérifiées afin d'identifier celles potentiellement frauduleuses. Une fois ces modèles mis en place, le choix des demandes à vérifier sera plus exact, générera davantage d'argent via les révisions à la baisse et permettra d'économiser du temps et de la main d'œuvre

Créez des modèles pour détecter les demandes d'indemnisation frauduleuses

Un grand prestataire d'assurance doit déterminer avec précision quelles demandes d'indemnisation sont frauduleuses afin de pouvoir s'employer à éviter toute perte de revenus. Au fil des ans, cette organisation a collecté les résultats des vérifications de demandes d'indemnisation. Elle n'a que rarement utilisé ses archives pour identifier les possibles demandes frauduleuses à venir. Ses méthodes précédentes lui faisaient souvent rater des occasions de récupérer de l'argent et ses experts passaient trop de temps à examiner des demandes légitimes.

Le data mining permet désormais à cette compagnie de prédire quelles demandes sont susceptibles d'être frauduleuses. Les experts peuvent ainsi déterminer quels dossiers cibler, récupérer des millions de dollars autrement perdus et gagner un temps précieux.

Le bureau de détection des fraudes de cette compagnie d'assurance a utilisé IBM SPSS Modeler, l'utilitaire de data mining de référence, pour obtenir des résultats. Modeler examine chaque ligne remplie sur les demandes d'indemnisation, compare les lignes remplies à la somme en dollars de la fraude détectée, classe les demandes par ordre de probabilité de fraude et affiche les résultats.

Le scénario suivant explique comment l'agence a créé des modèles et prédit quelles demandes risquaient d'être frauduleuses.

Interprétez vos données

L'interface de programmation visuelle d'IBM SPSS Modeler simplifie l'examen et la modélisation des dossiers de vérification. Les dossiers utilisés pour modéliser les demandes de remboursement de frais médicaux peuvent inclure les relevés de facturation d'assurance maladie avec des informations détaillées comme les codes et le pays de résidence du bénéficiaire/prestataire, les codes de diagnostic, la source d'admission, la durée du séjour et le montant total réclamé.

Déterminez la composition de votre population

Une étape importante du processus de data mining est de veiller continuellement à utiliser les bonnes données pour résoudre votre problème commercial. Vous devez vous assurer que les données ne représentent pas de façon disproportionnée un prestataire ou n'associent pas de façon exclusive un prestataire à un code de diagnostic particulier.

Découvrez des relations dans vos données

IBM SPSS Modeler ajoute facilement de nouvelles variables à chaque dossier de l'ensemble de données. Il est ensuite possible d'acheminer des données vers un nœud qui générera un graphique en toile d'araignée pour examiner à quelle fréquence (et pour quels codes de diagnostic) chaque prestataire a présenté des demandes d'indemnisation pour des services extérieurs à sa région. Ces informations pourraient s'avérer utiles plus loin dans notre analyse.

Créez un modèle

Dans cette étape, nous modélisons le montant total d'une demande d'indemnisation, en utilisant la source d'admission, la durée du séjour et les codes de diagnostic comme entrées. Nous avons choisi une procédure de modélisation appelée généralisation de règle parce qu'elle est facile à comprendre. Lorsque nous insérerons le modèle dans le flux, le modèle lira les entrées (source d'admission, durée du séjour et code de diagnostic) de chaque dossier, puis produira une valeur projetée pour le montant total. Nous utiliserons cette nouvelle valeur plus tard.

Comparez le modèle aux dossiers réels

Pour examiner la différence entre les frais réels consignés sur chaque demande d'indemnisation et les frais que notre modèle a projetés, nous les comparerons dans un diagramme de dispersion. Sur ce diagramme, nous pourrions ajouter un graphe de la ligne $y=x$. Si les frais réels sont égaux aux frais projetés sur un dossier particulier, le point de ce dossier doit se trouver sur cette ligne. En revanche, si les frais réels du dossier étaient supérieurs à ceux projetés par le modèle, son point se trouverait au-dessus de cette ligne.

Segmentez vos données

Pour explorer plus en détail les différences entre les frais réels et projetés, nous dérivons une nouvelle variable à ajouter à nos données. Cette nouvelle variable, « différentielle », est alors représentée dans un histogramme. On y observe que la majorité des dossiers dans nos données avait une valeur différentielle classifiée très proche de 0 \$. Toutefois, peu de dossiers ont une valeur différentielle dépassant 10 000 \$. Nous pouvons maintenant recentrer notre attention sur ces derniers dossiers.

Comparez votre sous-ensemble à l'ensemble de la population

Supposez que deux prestataires aient eu des demandes d'indemnisation d'une importance disproportionnée. Explorer les types de demandes soumises peut révéler une activité très suspecte de la part de l'un des demandeurs. Ils n'ont déposé des demandes que pour un code de diagnostic. L'analyste commercial peut facilement identifier ce comportement fortement suspect et étudier en détail les demandes d'indemnisation. Le data mining s'est avéré utile pour deux raisons. Premièrement, il a fourni de précieuses informations sur les cas de fraude potentiels dans les dossiers actuellement archivés. Bien qu'il ne s'agisse pas d'un cas flagrant de demande d'indemnisation frauduleuse, les preuves que nous avons rassemblées peuvent maintenant être transmises aux enquêteurs.

En outre, le modèle de data mining d'IBM SPSS que nous avons créé peut s'appliquer à de futures demandes. Il calculera les frais totaux projetés des demandes à venir. Ces projections pourront être comparées aux frais réels, et le système « signalera » les demandes douteuses pour qu'elles fassent l'objet d'une investigation. Grâce aux informations découvertes via le data mining, les experts peuvent se concentrer sur les demandes les plus susceptibles d'être revues à la baisse, et perdront

moins de temps à examiner des demandes légitimes. Avec le data mining, vos experts peuvent s'employer à récupérer de l'argent afin que les résultats de votre organisation soient moins affectés par la fraude.

Bien que le bureau de détection des fraudes de cette agence d'assurance ait utilisé le data mining pour la fraude des prestataires, vous pouvez également l'utiliser pour :

- la fraude à l'admissibilité
- la fraude à l'assurance automobile
- la fraude au crédit
- la fraude en ligne
- ainsi que beaucoup d'autres types de fraude

Et parce que les circonstances évoluent au fil du temps, vous pouvez revoir périodiquement les modèles et les actualiser afin qu'ils restent efficaces et offrent des résultats optimaux. IBM SPSS Modeler est le seul produit de data mining qui permet aux organisations de modifier continuellement le processus de data mining pour que les décideurs restent au courant des évolutions.

Déployez vos résultats de data mining de façon stratégique pour un succès optimal

Une fois que vous avez les modèles qui prédisent l'activité frauduleuse, vous devez déployer de façon stratégique vos résultats aux personnes qui peuvent utiliser ces informations pour éradiquer la fraude et récupérer de l'argent. Un déploiement stratégique signifie intégrer les modèles aux activités quotidiennes de votre entreprise. Un déploiement stratégique vous permet de mettre des informations cohérentes et opportunes dans les bonnes mains. Tous les membres de votre organisation disposent des mêmes informations et peuvent agir plus rapidement pour récupérer un maximum d'argent.

IBM® SPSS® Decision Management vous permet de comparer les valeurs aux nouvelles déclarations de sinistres et demandes d'indemnisation, puis de déployer votre modèle et de diffuser les résultats (par exemple, une liste des demandes les plus susceptibles d'être non conformes). Selon vos besoins, vous pouvez communiquer ces résultats sur un intranet, par e-mail ou sur des rapports papier.

Pour les entreprises qui ont des bureaux locaux ou des succursales, un déploiement stratégique de data mining apporte un avantage supplémentaire. Le bureau central peut stocker et explorer des données pour l'organisation entière et déployer les résultats de data mining vers les bureaux locaux, qui sont souvent chargés de combattre la fraude et les abus. Déployer les résultats de data mining vers les bureaux locaux peut étendre des ressources restreintes, et permettre ainsi de partager des informations de façon cohérente dans toute l'organisation.

Vous ne savez pas par où commencer ?

Vous pouvez commencer par suivre et résoudre vos problèmes commerciaux cruciaux en utilisant la meilleure technique de data mining, le Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM est une méthodologie de data mining complète et un modèle de processus qui rend les gros projets de data mining plus rapides, plus efficaces et moins coûteux. Notre compagnie adhère à cette approche optimale du data mining, dont plusieurs autres compagnies de premier plan sont co-auteurs, et la met au service de votre organisation pour vous offrir des résultats exploitables. Le modèle CRISP-DM offre des instructions pas-à-pas, des tâches et des objectifs pour chaque étape du processus, comme la compréhension de l'activité, la compréhension des données, la préparation des données, la modélisation, l'évaluation et le déploiement. Cette méthodologie peut constituer un excellent point de départ pour vos efforts de data mining en vous aidant à :

- évaluer et hiérarchiser les problèmes commerciaux
- élaborer des méthodes de data mining pour les résoudre
- appliquer des techniques de data mining
- interpréter les résultats de data mining
- déployer et tenir à jour les résultats de data mining

Pour plus d'informations sur CRISP-DM, voir www.crisp-dm.org.

Le data mining fait la différence

Découvrez des schémas qui indiquent quelles demandes sont les plus susceptibles d'être frauduleuses en appliquant des techniques sophistiquées de data mining à vos données sur les demandes d'indemnisation passées.

Les solutions d'IBM SPSS vous aident à mieux utiliser les données et la technologie et améliorent votre capacité à récupérer d'importantes sommes d'argent. Nos experts font équipe avec vous pour intégrer des connaissances analytiques et techniques approfondies à votre expertise métier. Présents à vos côtés, ils forment votre personnel, émettent des recommandations et créent un processus répétable afin que votre organisation puisse appliquer ces compétences et outils pour progresser facilement toute seule.

Analysez vos données à l'aide d'un ensemble de techniques, allant des simples rapports aux méthodes avancées permettant de prédire le comportement de vos prestataires, clients ou fournisseurs. Il est important d'utiliser plusieurs méthodes d'analyse pour pouvoir travailler avec de nombreux types de données dans de nombreuses applications et de toujours obtenir les réponses qui vous offrent les meilleures chances de revoir les sinistres à la baisse.

À propos d'IBM Business Analytics

Les logiciels IBM Business Analytics fournissent des informations complètes, cohérentes et précises permettant aux décideurs d'améliorer les performances de leur entreprise. Un portefeuille complet de solutions de Business Intelligence, d'Analyse Prédictive, de Performance financière & gestion de la stratégie et d'Applications Analytiques permettent d'avoir une vision claire et précise de la situation actuelle et de prédire les événements futurs. Combinées à de puissantes solutions métiers, les organisations de toute taille peuvent améliorer leur productivité, optimiser leur prise de décisions et délivrer de meilleurs résultats.

Dans ce cadre, les logiciels IBM SPSS Predictive Analytics aident les organisations à prévoir les événements futurs pour une meilleure prise de décisions et donc une amélioration de leurs résultats. Nos clients des secteurs privé, public et universitaire se fient à la technologie IBM SPSS et considèrent cette technologie comme un avantage concurrentiel pour : fidéliser sa clientèle, attirer de nouveaux clients, maximiser ses ventes, réduire la fraude, minimiser les risques. En intégrant le logiciel IBM SPSS à leurs opérations quotidiennes, les organisations deviennent des entreprises prédictives capables d'optimiser leur prise de décision afin d'atteindre leurs objectifs et d'obtenir un avantage concurrentiel important. Pour plus de renseignements, visitez le site www.ibm.com/spss/fr.



© Copyright IBM Corporation 2010

IBM Corporation
Route 100
Somers, NY 10589

Droits restreints pour les utilisateurs du gouvernement américain - l'utilisation, la duplication ou la divulgation sont soumises aux restrictions visées dans le contrat GSA ADP Schedule conclu avec IBM Corp.

Produit aux États-Unis d'Amérique
Mai 2010
Tous droits réservés

IBM, le logo d'IBM, ibm.com, WebSphere, InfoSphere et Cognos sont des marques commerciales ou des marques déposées d'International Business Machines Corporation aux États-Unis, dans d'autres pays ou les deux. Si ces termes ou si d'autres termes déposés d'IBM sont représentés pour la première fois dans ce document suivis du symbole de marque déposée (® ou TM), ces symboles indiquent des marques déposées ou de droit commun appartenant à IBM lors de la publication de ce document. Ces marques commerciales doivent également faire l'objet d'un enregistrement ou doivent être des marques de droit commun dans les autres pays. Une liste des marques commerciales actuelles d'IBM est disponible sur Internet sous « Droits d'auteur et marques » à l'adresse www.ibm.com/legal/copytrade.shtml.

SPSS est une marque commerciale de SPSS, Inc., an IBM Company, déposée dans de nombreuses juridictions dans le monde.

Les autres noms d'entreprises, de produits ou de services peuvent être des marques commerciales ou des marques de service d'autres organisations.



Réutilisez s'il vous plaît