

Analyse et application de modèles de régression pour optimiser le retour sur investissement d'opérations commerciales

Thomas Piton^{*,**}, Julien Blanchard^{**}, Henri Briand^{**}, Laurent Tessier^{***}, Gaëtan Blain^{*},

* Groupe VM Matériaux, Route de la Roche sur Yon, 85 260 L'Herbergement
{tpiton, gblain}@vm-materiaux.fr, <http://www.vm-materiaux.fr/>

** LINA équipe COD - UMR 6241 CNRS, 2 rue de la Houssinière, 44322 Nantes
{julien.blanchard, henri.briand}@univ-nantes.fr, <http://www.polytech.univ-nantes.fr/COD>

*** KXEN, 25 quai Galliéni, 92158 Suresnes
laurent.tessier@kxen.com, <http://www.kxen.com/>

VM Matériaux, entreprise de Négoce de matériaux, de menuiserie industrielle et de béton prêt à l'emploi réalise de nombreuses opérations commerciales, ciblant principalement ses clients professionnels. Pour une grande partie des campagnes, une invitation à participer est envoyée à chaque client « routé ». Le routage est réalisé manuellement par l'équipe marketing quelques semaines avant l'opération et se base principalement sur les clients ayant réalisé un certain seuil de chiffre d'affaire (CA) l'année précédente. Ces opérations commerciales maîtrisées depuis une dizaine d'années mettent en jeu des dépenses et des recettes importantes.

Dès lors, le retour sur investissement (*Return On Investment* ou ROI) des opérations commerciales de VM Matériaux peut être amélioré par des techniques de fouille de données. La connaissance extraite des différents modèles doit permettre aux experts de comprendre le comportement de leurs clients et ainsi prendre des décisions en utilisant le savoir extrait à bon escient (paradigme de l'actionable knowledge (Cao, 2007; Graco et al., 2007)). Le retour d'expérience du projet de fouille de données mené chez VM Matériaux pour améliorer le ROI des opérations commerciales a été positif. Nous avons développé plus particulièrement l'évaluation et la mise en œuvre des modèles de régression *ridge* (Dodge, 2004) pour perfectionner le routage d'une campagne marketing. Ces modèles ont été construits avec le logiciel KXEN qui se fonde sur la théorie de l'apprentissage statistique (Vapnik, 1998).

À l'aide de l'entrepôt de données existant, nous avons créé un modèle de données basé sur les clients routés l'année précédente. Nous avons ajouté leurs caractéristiques de la table des clients. Ensuite, nous avons enrichi le modèle avec le résultat d'une opération commerciale similaire mais printanière. Par la suite, nous avons créé des agrégats temporels basés sur le chiffre d'affaire, la marge nette et le nombre de lignes de commandes sur six périodes de un mois. Enfin, nous avons ajouté une cible binaire relative à la détection des acheteurs (égale à 1 si le client a acheté, à 0 sinon). Cette phase de pré-traitement des données génère un modèle de 66 variables et de 10 378 lignes.

L'application d'un jeu de données a permis de générer un score et une probabilité pour chaque client, soulignant le potentiel que chacun achète ou non durant l'opération commerciale. L'ensemble des clients a été trié par probabilité de participation.