

Tirer profit des sources externes pour l'enrichissement des bases clients

Application du Data Mining prédictif aux bases EDF

Christian Derquenne*, Sabine Goutier*, Sylvia Lembo** et Véronique Stéphan*

* Electricité de France, Recherche et Développement

Département ICAME

1, av. du Général de Gaulle, 92 141 Clamart Cedex, France

{christian.derquenne, sabine.goutier, veronique.stephan}@edf.fr

** Ardans SAS

Le Cristal – 2, rue Hélène Boucher, 78286 Guyancourt Cedex, France
slembo@ardans.fr

Résumé. L'objet de l'article est de montrer comment l'utilisation de techniques prédictives de Data Mining améliorent la qualité de l'information utilisée en marketing opérationnel. Cette étude s'inscrit dans la continuité des méthodologies utilisées à EDF pour enrichir les bases clientèles à partir de modèles prédictifs construits sur des variables internes. Ces modèles sont généralement construits à partir de variables explicatives issues des bases de données clients. Afin d'améliorer la qualité des modèles, nous prenons en compte des données externes agrégées. Nous présentons dans cet article comment utiliser conjointement des variables explicatives issues des bases de données clients et l'information supplémentaire issue du recensement de la population de 1999. La méthodologie a été expérimentée sur des bases clients EDF dans le cadre de campagnes de marketing opérationnel.

1. Contexte

Notre travail s'inscrit dans le cadre des techniques prédictives de Data Mining appliquées à EDF pour le segment de clientèle Particuliers. L'utilisation de ces techniques visent à améliorer la connaissance des clients de l'Entreprise permettant ainsi une meilleure efficacité des campagnes de marketing opérationnel. Les variables utilisées pour le ciblage sont en effet très inégalement renseignées dans les bases de données. Pour s'affranchir des valeurs manquantes, une méthodologie d'enrichissement des bases de données clients à l'aide de modèles de prédiction a été définie et expérimentée au sein de l'Entreprise (Derquenne 2000). Les caractéristiques disponibles sur seulement une partie de la clientèle sont ainsi extrapolées à l'ensemble des clients.

Lors de la mise en place d'une campagne de marketing direct, les critères permettant de cibler les clients visés par cette campagne sont ainsi mieux qualifiés, ce qui permet d'augmenter la taille de la cible et d'améliorer le rendement de la campagne.

L'approche proposée dans cet article complète la démarche de Data Mining initiale en intégrant aux modèles de prédiction des nouvelles variables explicatives construites à partir d'informations externes issues du recensement de la population de 1999 (RP99). En effet, depuis ces dernières années, les techniques de ciblage s'affinent prenant également en