

**MISE EN ŒUVRE D'UNE DEMARCHE STATISTIQUE COMPLETE**  
**POUR LA PREDICTION DE VARIABLES**  
**DANS UNE BASE DE DONNEES CLIENTELE D'EDF**

**Christian Derquenne**

*EDF – Division Recherche et Développement  
1 avenue du Général de Gaulle  
92141 CLAMART Cedex*

*Cet article fait partie des Actes des 5<sup>èmes</sup> JOURNEES MODULAD qui ont eu lieu les 16 et 17 novembre 2000 à E.D.F. (Clamart).*

## **1. Contexte, historique et objectif**

Mieux connaître les attentes et les besoins de la clientèle est un des axes stratégiques majeurs d'EDF. Cette meilleure connaissance est notamment obtenue grâce à l'analyse statistique de différents types de données (consommation d'électricité, fidélisation, nouveaux clients, qualité de fourniture, enquêtes de satisfaction, ...). Pour cela, EDF possède plusieurs bases de données (fichiers de facturation, enquêtes d'opinion, ...) sur différents segments de clientèle (résidentielle, professionnelle, PME-PMI, grandes entreprises). L'objectif des études peut être par exemple de prédire l'énergie de chauffage ou d'identifier les clients mécontents.

La mise en œuvre de ce type d'analyse a fait l'objet d'un projet qui a débuté en 1998 à la Division Recherche et Développement d'EDF. A l'origine, ce projet avait pour but d'évaluer la faisabilité et l'intérêt de l'application des techniques de Data Mining sur les différentes données de la clientèle d'EDF. En résultat du projet, on attendait par conséquent, des