

Classification supervisée

Travaux pratiques sous le logiciel R

Nous allons travailler sur une base de données décrivant les clients d'une banque et leurs comportements (mouvements, soldes des différents comptes) :

<http://eric.univ-lyon2.fr/jjacques/Download/DataSet/VisaPremier.txt>

L'objectif final est l'estimation d'un score d'appétence à la carte VISA Premier. C'est une carte de paiement haut de gamme qui cherche à renforcer le lien de proximité avec la banque en vue de fidéliser une clientèle aisée.

Après échantillonnage et une première sélection de variables, la base de données (VisaPremier.txt) est composée de 1073 clients décrits sur les variables suivantes :

Identif.	Libellé	Identif.	Libellé
matricul	Matricule (identifiant client)	mtfactur	Montant facturé dans l'année en francs
departem	Département de résidence	engageml	Engagement long terme
ptvente	Point de vente	nbvie	Nombre de produits contrats vie
sexe	Sexe (qualitatif)	mtvie	Montant des produits contrats vie en francs
age	Age en année	nbeparmo	Nombre de produits épargne monétaire
sitfamil	Situation familiale (Fmar : marié, Fcel : célibataire, Fdiv : divorcé, Fuli : union libre, Fsep : séparé de corps, Fveu : veuf)	mteparmo	Montant des produits d'épargne monétaire en francs
ancienne	Ancienneté de relation en mois	nbeparlo	Nombre de produits d'épargne logement
csp	Catégorie socio-professionnelle (code num)	mteparlo	Montant des produits d'épargne logement en francs
codeqlt	Code "qualité" client évalué par la banque	nbilivret	Nombre de comptes sur livret
nbimpaye	Nombre d'impayés en cours	mtlivret	Montant des comptes sur livret en francs
mtrejet	Montant total des rejets en francs	nbeparlt	Nombre de produits d'épargne long terme
nbopguic	Nombre d'opérations par guichet dans le mois	mteparlt	Montant des produits d'épargne long terme en francs
moycred3	Moyenne des mouvements nets créditeurs des 3 mois en kF	nbeparte	Nombre de produits épargne à terme
aveparmo	Total des avoirs épargne monétaire en francs	mtparate	Montant des produits épargne à terme
endette	Taux d'endettement	nbbon	Nombre de produits bons et certificats
engagemt	Total des engagements en francs	mtbon	Montant des produits bons et certificats en francs
engagemc	Total des engagements court terme en francs	nbpaiecb	Nombre de paiements par carte bancaire à M-1
engagemm	Total des engagements moyen terme en francs	nbcb	Nombre total de cartes
nbcptvue	Nombre de comptes à vue	nbcbptar	Nombre de cartes point argent
moysold3	Moyenne des soldes moyens sur 3 mois	avtscpte	Total des avoirs sur tous les comptes
moycredi	Moyenne des mouvements créditeurs en kF	aveparfi	Total des avoirs épargne financière en francs
agemvt	Age du dernier mouvement (en jours)	cartevp	Possession de la carte Visa Premier
nbop	Nombre d'opérations à M-1	sexer	Sexe codé en 0/1
		cartevpr	Possession de la carte Visa Premier codé en 0/1
		nbjdebit	Nombre de jours de débit

1. Identifier les variables quantitatives et les variables catégorielles. Réaliser une analyse descriptive univariée sommaire des variables à l'aide de la fonction `summary`, ce qui vous permettra d'identifier certains soucis potentiels dans les données (erreur de codage, variables redondantes, variables constantes...)
2. Réaliser une ACP sur les variables quantitatives. Représenter les individus.
3. Réaliser une classification pour prédire l'achat de la carte Visa Premier. Évaluer votre modèle à l'aide d'un échantillon test que vous aurez au préalable extrait de l'échantillon total.