

# MODELISATION DU TEMPS DE REPONSE A UNE DEMANDE DE CREDIT: CAS DE AFRILAND FIRST BANK.

Par WOMGA Albert Tchaptchet

Master de Statistique Appliquée  
ENSP - Université de Yaoundé 1

24 Octobre 2007

# Généralités

L'activité principale de la banque consiste à exercer le rôle d'intermédiaire financier entre des personnes physiques ou morales à excédent de capitaux et ceux qui ont un besoin en financement.

## source de financement de la banque

- Collecte des fonds des épargnants
- Prêt auprès des grandes institutions.

# Généralités

## Conséquences :

Les banques sont soumises à des normes qui dérogent aux dispositions du droit commercial et du droit civil: Fond propres de la banque  $> 8\%$  des (risques de crédits (75%) + de marché (5%) + opérationnels (20%)).

## solution

Système d'octroi collégiale

# Généralités

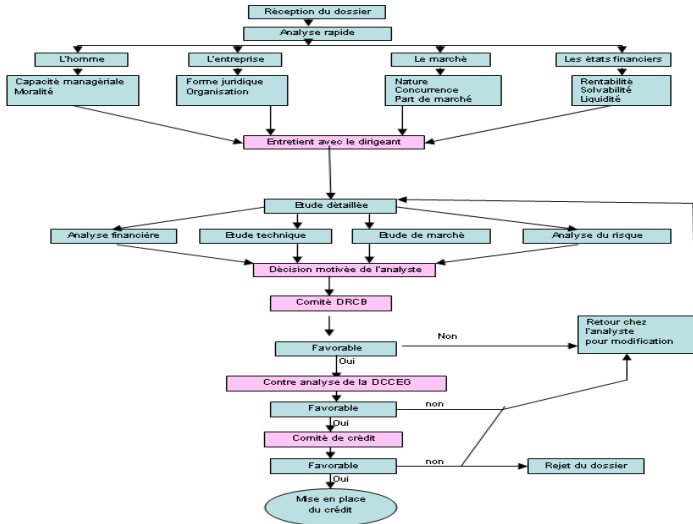
## Conséquences :

Les banques sont soumises à des normes qui dérogent aux dispositions du droit commercial et du droit civil: Fond propres de la banque  $> 8\%$  des (risques de crédits (75%) + de marché (5%) + opérationnels (20%)).

## solution

Système d'octroi collégiale

## Procédure



# Contexte

- Près de (67%) de réclamation
- La rigidité de l'analyste à accepter ce dernier

# Problématique

Comment optimiser le temps le temps d'analyse d'un dossier de crédit?

## Plan de l'exposé

- I- Analyse descriptive
- II- Objectifs
- III- Méthode utilisées
- IV- Résultats
- V- Conclusion

## Description des données : état d'avancement des dossiers de crédit

- Source : portefeuille du gestionnaire de crédit.
- enregistrements pour les exercices 2004-2006.
- Nombre de dossiers 78.
- 19 variables.

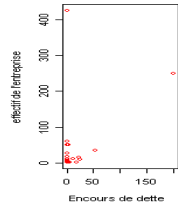
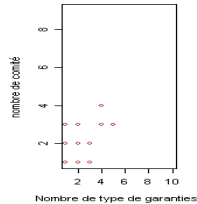
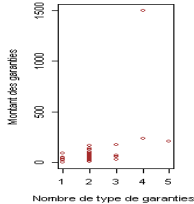
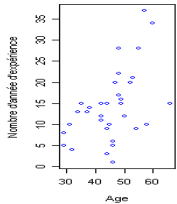
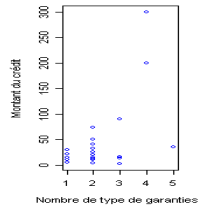
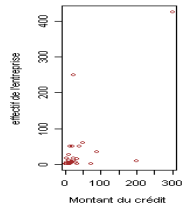
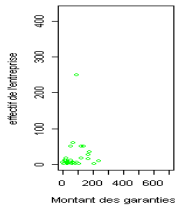
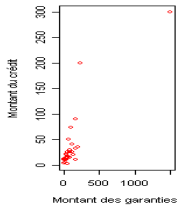
# Analyse des liaisons entre les variables quantitatives et le temps

	mc.t	ee.t	mcr.t	mg.t	nvg.t	ag.t	ex.t
mc.t	1	-0,07	-0,06	-0,03	-0,02	0,08	-0,06
ee.t	-0,07	1	0,69	0,84	0,27	0,17	0,11
mcr.t	-0,06	0,69	1	0,87	0,54	0,08	-0,02
mg.t	-0,03	0,84	0,87	1	0,48	0,14	0,02
nvg.t	-0,02	0,27	0,54	0,48	1	-0,21	-0,23
ag.t	0,08	0,17	0,08	0,14	-0,21	1	0,52
ex.t	-0,06	0,11	-0,02	0,02	-0,23	0,52	1
det.t	-0,05	0,43	0,08	0	0,06	0,1	-0,04
co.t	0,01	0,25	0,62	0,4	0,48	-0,2	-0,28
mvc.t	0,24	-0,24	-0,2	-0,22	-0,03	0,06	0,13

## Analyse des liaisons entre les variables quantitatives et le temps (suite)

	dr.t	da.t	dn.t	dsc.t	dm.t
mc.t	0,3	-0,07	-0,15	-0,19	-0,14
ee.t	0,22	-0,07	-0,01	-0,06	0,1
mcr.t	-0,09	-0,13	-0,16	-0,09	-0,01
mg.t	-0,03	-0,09	-0,05	-0,07	0,09
nvg.t	-0,36	-0,1	-0,2	-0,1	-0,01
ag.t	0,28	-0,16	-0,06	-0,1	0,02
ex.t	0,35	0,09	0,08	0,08	-0,06
det.t	0,3	-0,15	-0,08	-0,23	0,04
co.t	-0,21	-0	-0,34	-0,14	-0,23
mvc.t	0,1	0,03	0	-0,05	-0,03
dr.t	1	-0,07	0,04	-0,15	0,13

# Analyse des liaisons entre les covariables quantitatives

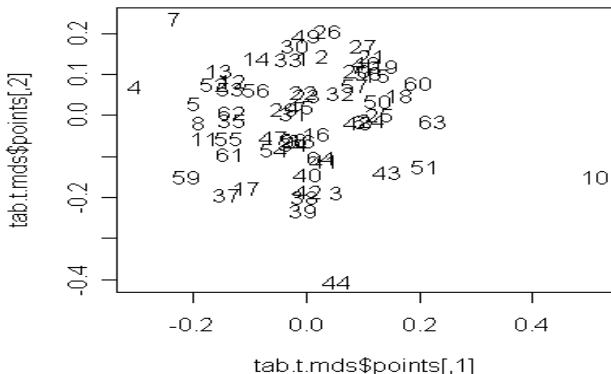


## Analyse des liaisons entre les variables qualitatives et le temps (suite)

variables	rap,cor
da & fj	0,06
da & tc	0,02
da & nat	0,09
da & le	0,13
da & ev	0,17

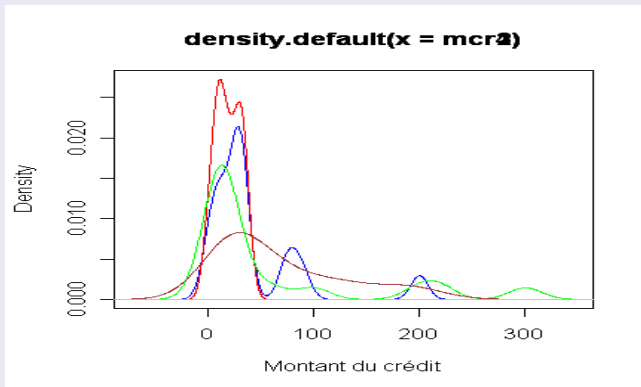
## Positionnement multidimensionnel

Ici on projet les individus dans un même plan afin de déceler s'il existe des groupes qui se distinguent. Au regard du graphe ci-dessous, on n'arrive pas à visualiser des groupements pertinents.



# Epuration du jeu de données

## Etude des comportements des variables suivant les classes



## Epuration du jeu de données suite

### Etude des comportements des variables suivant les classes suite

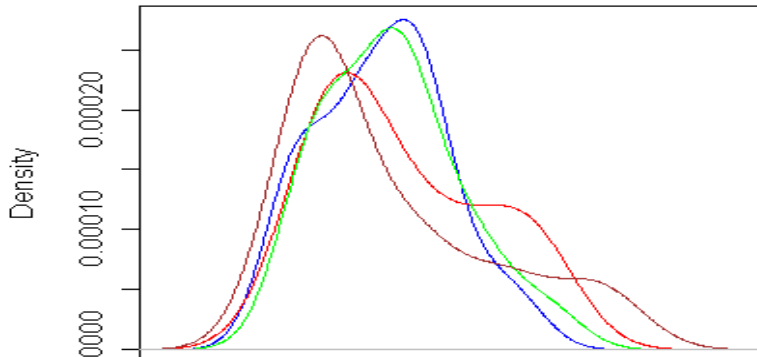
Groupe	p-value
4 & 2	0,03109
4 & 3	0,02002

Table: Montant du crédit

## Epuration du jeu de données (suite)

### Etude des comportements des variables suivant les classes (suite)

**density.default(x = dr2)**



## Epuration du jeu de données (suite)

### Etude des comportements des variables suivant les classes (suite)

Groupe	p-value
1 & 2	0,6496
1 & 3	0,527
1 & 4	0,6113
2 & 3	0,708
2 & 4	0,7903
3 & 4	0,5684

# Objectif

vu la difficulté à décélérer une éventuelle relation avec la variable à expliquer pouvoir regroupe les individus suivant les classes selection plus haut afin de posséder à une discretisation

## Méthode k-NN

### dissimilarité

- si  $X_j$  est quantitative:

$$d_j(x_j^i, x_j^{i'}) = \begin{cases} |x_j^i - x_j^{i'}| \\ \text{ou} \\ (x_j^i - x_j^{i'})^2 \end{cases}$$

et l'on choisira la distance qui donne ma meilleur prédiction.  
En général, on utilise la métrique euclidienne.

- si  $X_j$  est qualitative:  $d_j(x_j^i, x_j^{i'}) = \mathbb{I}_{x_j^i \neq x_j^{i'}}$ .

On définit alors  $\forall i, i' \in \mathcal{A}$ ,  $d(i, i') = \sum_{j=1}^p d_j(x_j^i, x_j^{i'})$ .

# Méthode CART

## Construction de la séquence d'arbres

Pour un arbre  $A$  donné, on note  $K$  le nombre de feuilles ou nœuds terminaux de  $A$ ; la valeur de  $K$  exprime la complexité de  $A$ . La mesure de la qualité de discrimination d'un arbre  $A$  s'exprime par un critère

$$D(A) = \sum_{k=1}^K D_k(A),$$

où  $D_k(A)$  est le nombre de mal classés ou la déviance ou le coût de mauvais classement de la  $k$ ème feuille de l'arbre  $A$ .

La construction de la séquence d'arbres emboîtés repose sur une pénalisation de la complexité de l'arbre:

$$C(A) = D(A) + \gamma K.$$

## Méthode k-NN

Le nombre de voisin est de 6, avec un taux d'erreur de prévision par validation croisé (leave-one-out) au tour de 45%. L'on constate pour un nouveau dossier le prédicteur k-NN renvoie la classe dominante parmi les classes de ses voisins 6 les plus proches. Ainsi la durée d'analyse vaut la moyenne des durées d'analyse des individus qui tombe dans cette classe.

La construction de l'arbre a nécessité au préalable une détermination du coefficient de complexité qui est de 0.004. Les variables qui ont contribué à la mise en place de cet arbre sont : nat, comp1, comp2, comp3, comp4. Les feuilles sont au nombre de 7.

Classe	Probabilité à priori	Probabilité à posteriori
1	0,22	0,285
2	0,22	0,285
3	0,35	0,285
4	0,22	0,285

## Conclusion

- Avec la méthode des k plus proche avec  $k = 6$  voisin on obtenu un taux d'erreur au tour de 45% ceci était prévisible puisqu'on ne disposait pas assez d'information pour discriminer entre les dossiers étant données la formation des groupes n'était pas perceptible. On a enregistré également un taux de erreur acceptable au tour de 15% pour la méthode des arbres de prédiction. En égard à ce qui précède, le meilleur modèle serait la méthode des arbres de prédiction.

## Prespectives

- L'étude donc nous avons conduit pour ce mémoire nécessite des améliorations à plusieurs niveaux. Entre autres je citerai : la qualité du prédicteur (temps machine important voir des heures), le prédicteur ne répondant pas objectivement à la dite préoccupation.
- le modèle de suivi serait mieux si l'on avait l'information sur les clients qui n'ont pas eu de suite favorable.

## Prespectives

- L'étude donc nous avons conduit pour ce mémoire nécessite des améliorations à plusieurs niveaux. Entre autres je citerai : la qualité du prédicteur (temps machine important voir des heures), le prédicteur ne répondant pas objectivement à la dite préoccupation.
- le modèle de suivi serait mieux si l'on avait l'information sur les clients qui n'ont pas eu de suite favorable.

MERCI POUR VOTRE  
AIMABLE  
ATTENTION.