

Prédiction de la défaillance de la clientèle du micro crédit au Maroc : méthodes des scores

Par Omar TAOUAB¹

Résumé

Le développement de la micro finance au Maroc date du milieu des années 1990 et devient particulièrement dynamique au cours des années 2000. Actuellement, ce secteur commence à connaître de multiples risques notamment celui du crédit.

En effet, le risque de crédit, susceptible d'entraver l'émergence des établissements de micro-finance, est devenu l'une des préoccupations principales des institutions financières. Ainsi, une place importante mérite d'être réservée au métier de la gestion des risques d'insolvabilité des emprunteurs notamment à travers l'anticipation de leurs défaillances .Cet objectif d'anticipation est réalisable à travers l'utilisation d'une technique de scores qui permet d'affecter un individu en l'un des groupes à savoir les bons clients et les défaillants.

En s'inspirant des enseignements sur la fonction de discrimination présentés par (SAPORTA,1975), le discernement peut être effectué selon les caractéristiques socio-économiques et démographiques de chaque client et compte tenu des modalités prises par les différentes variables.

Le calcul des scores nous a permis élaborer un filtre de discrimination permettant de noter les clients, d'identifier des classes de risque et de repérer ceux susceptibles de tomber défaillants à la date de remboursement.

Mots clés : Micro crédit, Risque, Défaillance, Scoring, , résultats, propositions.

¹ Enseignant chercheur à L'ENCG, Kénitra- Maroc

Introduction

Les établissements de micro-credit commencent récemment à prendre d'envergure à côté des organismes spécialisés partout dans le monde. Au Maroc, ce développement intense confère à certain acteur tel que l'association Al amana un rôle non négligeable en matière d'octroi des crédits aux petits emprunteurs.

Toutefois, cette activité, qui ne cesse de s'élargir, n'est pas à exercer sans exposer l'association à des risques divers de la micro-finance et plus particulièrement la défaillance des clients. Dans ce contexte, l'adoption d'un système de détection de l'insolvabilité s'impose avec acuité pour pouvoir effectuer une catégorisation des clients en bon et mauvais.

Ce discernement peut être réalisé moyennant des scores établis selon les caractéristiques socio-économiques et démographiques de chaque client et compte tenu des modalités prises par les différentes variables. Ces scores nous permettent d'élaborer un filtre de discrimination permettant d'identifier les classes de risque et de repérer les clients susceptibles de tomber défaillants à la date de remboursement.

Pour mener à bien cet essai de recherche, nous proposons de présenter ce papier comme suit :

La première section passe en revue certains présupposés théoriques pour cerner les variables usuelles de la défaillance de la clientèle du micro-crédit. Une deuxième section est dédiée à la présentation de la méthodologie de recherche spécifique à notre problématique et la troisième section relate les résultats de l'estimation empirique assortis d'analyse pour conclure avec des propositions.

I- QUELQUES ETUDES ANTERIEURES :

Plusieurs études ont été menées sur l'analyse du risque de défaillance des demandeurs de crédit en utilisant la méthode des scores. On cite, à

titre d'exemple, celles faites par le groupe ACCION en Bolivie et en Equateur durant l'année 2002, et celles menées en Colombie et en République Dominicaine pendant l'année 2004 par Women's World Banking (Schreiner, 2007)¹.

L'objectif primordial de ces études était de vérifier que quelques informations et indicateurs déduits des dossiers relatifs aux clients suffisaient pour prédire la défaillance de ces derniers. Après une recherche théorique, nous allons nous inspirer d'une étude menée au Mali, sur l'institution de micro-finance de Nyèsigiso, qui a été réalisée par Boubacar Diallo en 2006, c'est la seule étude pour laquelle le rapport méthodologique s'est avéré plus détaillé et plus intéressant. Cette étude avait pour but, le développement d'un modèle de Crédit « Scoring » pour cette institution de micro-finance, afin de pouvoir qualifier l'éligibilité du client au financement.

II- Méthodologie

La méthodologie consiste à présenter d'une part, une présentation de la méthode des scores et d'autre part, les différentes variables explicatives ainsi que les différents tests pour aboutir sur des résultats et quelques propositions

1-La méthode des scores : Disqual

La méthode Disqual² (Discrimination sur variables qualitatives), introduite en 1975 par SAPORTA, consiste à :

- (i) effectuer, à partir de variables qualitatives quantitatives, judicieusement découpées en classes, une analyse des correspondances multiples (ACM) sur le tableau disjonctif complet de ces variables.

¹Mark Schreiner, le scoring peut-il attirer les investisseurs soucieux du profit vers le microcrédit?

² Diallo B. (2006) « Un modèle de «Credit scoring pour une institution de microfinance Africaine, le cas de Nyesigiso au Mali

- (ii) placer les coordonnées (continues) des individus sur les axes factoriels les plus discriminants issus de l'ACM,
- (iii) injecter ces coordonnées en entrée d'une analyse discriminante linéaire classique (AFD).et
- (iv) procéder à un changement de variables dans la fonction de score de Fisher pour l'exprimer en fonction des indicatrices des modalités des variables initiales (qualitatives).

Le résultat est donc une combinaison linéaire d'indicatrices de modalités, ce qui revient à attribuer une note à chacune de ces modalités.

III -Données

1-Préparation des données

Pour tout client sollicitant un micro-crédit auprès d'Al Amana, on trouve un formulaire de demande, sur lequel, figurent des informations telles que : l'âge du client, son sexe, son état matrimoniale, son milieu de résidence, le secteur d'activité où il exerce son métier, le montant de crédit sollicité, la fréquence de remboursement exprimée en nombre jours... Ces informations sont stockées simultanément et d'une manière automatique sur une base de données sur le serveur de l'association.

Pour des raisons de confidentialité, les données nous sont fournies en fonction des besoins du projet. La grande partie qui concerne les bons clients nous a été transmise sur des fichiers texte (Bloc-notes), et l'autre partie des mauvais clients nous a été confiée sur un fichier EXCEL.

Cependant, il existe quelques variables que nous avons jugées adéquates pour notre étude et qui ne sont pas disponibles, comme le montant des conditions financières (les frais du dossier et les autres frais similaire) relatif à chaque client, le revenu mensuel estimé de chaque client et la variable indiquant si un client a été inscrit, ou non, chez une autre institution exerçant dans le même secteur d'activité.

Nous avons converti ces fichiers sous format EXCEL pour aborder l'apurement des données disponibles et pour calculer d'autres variables à partir des premières.

La deuxième étape était le traitement des données, à l'aide de SPSS, afin de ne retenir que les variables nécessaires à l'étude et obtenir un fichier plus réduit et plus facile à exploiter, tout en contenant le maximum d'informations pertinentes. Le fichier en question fera objet d'une étude descriptive et sera la base de l'élaboration du modèle de « Scoring ». Cette étape est d'une grande importance car les résultats qui en émanent constituent la base de la suite de l'analyse.

Il en résulte un nouveau fichier contenant toutes les variables retenues et les variables créées, en respectant le fait que chaque observation ne figure qu'une et une seule fois dans la base de données.

2- Echantillonnage¹

Pour l'interrogation de la base de données, Al Amana utilise le Business Object. Le problème de cet outil de décision, c'est qu'il ne peut tirer qu'un nombre fini d'observations, donc, on n'a pas eu recours à toutes les observations figurant dans la base, afin d'extraire un échantillon en utilisant les techniques d'échantillonnage fondamentales pour effectuer le bon choix. Et pour ne pas influencer les résultats, nous avons décidé de travailler sur le maximum d'observations que le Business Object peut nous fournir.

Notre échantillon de travail sur lequel nous allons appliquer les méthodes d'analyse de la défaillance est composé de 84 391 observations. Et le tableau suivant, présente la répartition des bons et des mauvais clients dans l'échantillon.

¹ Fatima Bakass : Méthodologie des enquêtes statistiques, INSEA, 2007.

Tableau -1- : Répartition des bons et des mauvais clients dans l'échantillon de l'étude

	localisation		Total
	Urbain	rural	
Défaillants en %	16,33	6,09	22,36
Non défaillants en %	46,16	31,48	77,64
Total/ zone	62,49	37,51	100

IV-Testes statistiques

Les testes statistiques porteront d'une part sur la dépendance, la colinéarité des variables et d'autre part sur les coefficients du modèle et leur validation.

(i)Test de dépendance

Pour tester la dépendance entre la variable expliquée et les variables explicatives, nous utiliserons le test de khi-deux ou la matrice de corrélation.

L'hypothèse nulle est la suivante :

H_0 : il y a indépendance entre la variable expliquée Y et la variable explicative Xi.

Si la valeur calculée du Khi2 est supérieur à celle tabulée l'hypothèse sera rejetée.

(ii)Test de Kruscal Wallis

Pour tester l'indépendance entre les classes des variables explicatives, nous effectuerons le test de Kruksal Wallis puisqu'il ne suppose aucune hypothèse sur la loi de la variable de discrimination.

Lorsque tous les seuils critiques sont inférieurs à $\alpha = 5\%$, on rejette l'hypothèse nulle du test de Kruksal Wallis H_0 pour toutes les variables explicatives testées. Ce qui signifie que les classes formées sont différentes deux à deux.

(iii) Corrélation entre les variables explicatives

L'intensité de la liaison linéaire entre deux variables peut être mesurée par le coefficient de corrélation linéaire (ou R de Pearson). Alors, la liaison est :

- Nulle, si le coefficient de corrélation est nul ;
- Parfaite, si le coefficient de corrélation = +1 ou -1 ;
- Forte, si le coefficient de corrélation linéaire $> +0.7$ ou < -0.7 .

Le coefficient de corrélation linéaire est positif lorsque les deux variables évoluent dans le même sens : les deux augmentent ou diminuent ensemble. Un coefficient de corrélation négatif indique une variation inverse : l'une augmente quand l'autre diminue.

IV- Validation du modèle

La construction de l'outil de décision (fonction score) se fait toujours sur un échantillon dit échantillon d'apprentissage ou échantillon de base. Afin de savoir la qualité du modèle utilisé, plusieurs critères de performances peuvent être appliqués sur un échantillon test d'une taille suffisamment grande.

Parmi ces critères de performances que nous avons utilisées : les taux de bons classements, la courbe de ROC, l'indice de Gini et les boîtes à moustaches.

III- Résultats

Les principaux résultats sont déclinés, ci-après, conformément aux étapes présentées par la méthodologie

1-Résultats de la manipulation des données

Après une étude documentaire sur le micro-crédit et sur les problèmes rencontrés par les exerçants dans ce secteur, nous avons retenu une série de variables qui explique le défaut de paiement chez les clients des institutions de micro-finance. Or, comme nous l'avons déjà signalé, la base de données d'Al Amana ne permet pas de les prendre en compte dans leur totalité.

Les variables retenues pour le traitement des données sont :

- **nct** : le numéro de contrat ;
- **zone** : la zone géographique ;
- **local** : donne le milieu de résidence du client (Urbain/Rural) ;
- **sexe** : le sexe du client ;
- **datnais** : la date de naissance du client ;
- **etatmatr** : désigne l'état matrimonial du client ;
- **sectact** : indique le secteur d'activité du client ;
- **typpret** : montre s'il s'agit d'un prêt solidaire (PS) ou d'un prêt individuel (PI) ;
- **datderech** : la date de la dernière échéance ;
- **datdempre** : la date de demande du prêt ;
- **datobtpre** : la date d'obtention du prêt ;
- **mntdem** : indique le montant de prêt que le client a demandé (en DH) ;
- **mntacor** : indique le montant de prêt accordé au client (en DH) ;
- **period** : désigne la périodicité de remboursement.

A ces variables s'ajoutent de nouvelles variables calculées à partir des premières :

- **age** : l'âge du client ;
- **durpret** : indique la durée du prêt en nombre de jours, elle est égale à la différence entre la date de la dernière échéance et la date d'obtention du prêt ;
- **tempatt** : définit le temps d'attente pour l'obtention du prêt en nombre de jours, elle est égale à la différence entre la date d'obtention du prêt et la date de sa demande ;
- **degratio** : désigne le degré de rationnement, il est égale au rapport de la différence entre le montant de prêt demandé et le montant de prêt accordé, et le montant de prêt accordé.

Au final, les variables explicatives retenues pour analyser la défaillance sont : zone, local, sexe, etatmatr, sectact, typpret, mntacor, period, age, durpret, tempatt et degratio.

la variable dépendante (à expliquer) est la défaillance du client. Il s'agit d'une variable binaire notée « **default** » tel que :

- **default** =1 si le client est défaillant selon la définition adoptée par Al Amana ;
- **default** =0 si le client est sain selon la définition adoptée par Al Amana.

2- Résultats des tests statistiques

Ces résultats concernent aussi bien les données que la validation de la fonction de discernement (fonction des scores).

(i)Test de dépendance

Les résultats du teste khi-deux, présentés sur les annexes résultats , montre que pour toutes les variables, l'hypothèse d'indépendance est rejetée au seuil de signification 5%. Donc on peut dire qu'il y a une

dépendance entre la variable expliquée et les variables explicatives choisies.

A fin de s'assurer que les classes des variables sont significativement différentes, on a opté pour un test d'égalités des moyennes de ces classes.

(ii) Test de Kruscal Wallis

A partir des résultats de ce test, obtenu par SPSS portés sur les annexes résultats. On remarque que tous les seuils critiques sont inférieurs à $\alpha=5\%$, donc on rejette l'hypothèse nulle du test de Kruksal Wallis H_0 pour toutes les variables explicatives testées. Ce qui signifie que les classes formées sont différentes deux à deux.

(iii) Corrélation entre les variables

La matrice de corrélation entre les variables explicatives est donnée par le tableau suivant :

Tableau 2 : Matrice des corrélations.

	degratio	etatmatr	tempatt	typpret	sexe	period	zone	durpret	age	local	sectact	mmtacor
degratio	1	0,001	0,000	-0,003	0,040	0,007	0,000	-0,048	-0,009	0,029	-0,010	0,425
etatmatr	0,001	1	0,007	0,004	0,004	0,001	0,020	-0,009	0,263	0,036	-0,030	0,004
tempatt	0,000	0,007	1	-0,002	0,017	0,021	0,003	0,029	0,015	0,011	-0,003	-0,011
typpret	-0,003	0,004	-0,002	1	0,019	0,019	0,024	-0,010	0,004	-0,014	-0,011	0,011
sexe	0,040	0,004	0,017	0,019	1	0,002	0,009	-0,028	-0,048	0,153	-0,221	0,168
period	0,007	0,001	0,021	0,019	0,002	1	0,060	-0,127	-0,003	-0,083	0,018	0,049
zone	0,000	0,020	0,003	0,024	0,009	0,060	1	-0,081	-0,014	-0,138	0,102	0,011
durpret	-0,048	-0,009	0,029	-0,010	-0,028	-0,127	-0,081	1	-0,010	-0,065	0,076	-0,178
age	-0,009	0,263	0,015	-0,004	-0,048	-0,003	-0,014	-0,010	1	0,005	0,063	-0,056
local	0,029	0,036	0,011	-0,014	0,153	-0,083	-0,138	-0,065	0,005	1	0,319	0,121
sectact	-0,010	-0,030	-0,003	-0,011	-0,221	0,018	0,102	0,076	0,063	0,319	1	-0,080
mmtacor	0,425	0,004	-0,011	0,011	0,168	0,049	0,011	-0,178	-0,056	0,121	-0,080	1

On remarque que toutes les variables explicatives sont faiblement corrélées deux à deux, puisque tous les coefficients de corrélation sont inférieurs à 0,7. On remarque aussi, que la plus grande corrélation est

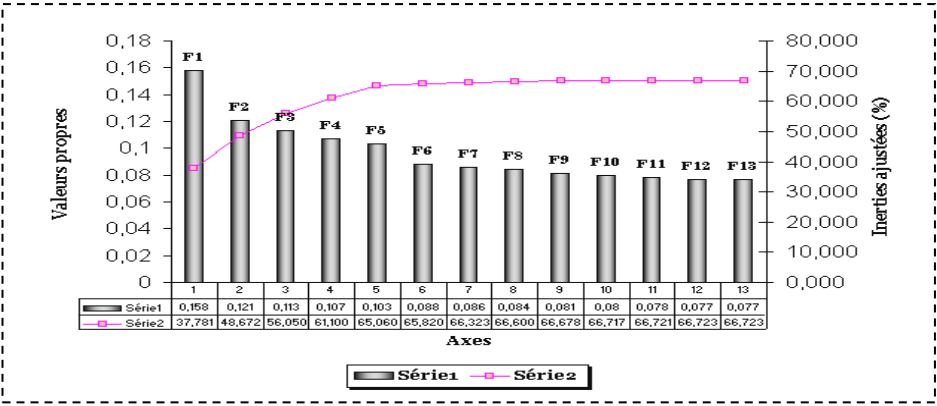
celle entre le mntacor (Montant accordé) et degratio (le degré de rationnement), avec un coefficient égale à 0,425.

3-Résultats de l’ACM

Pour appliquer l’ACM, nous avons utilisé le logiciel XLSTAT qui nous a permis d’avoir plus de résultats que le logiciel SPSS.

La figure suivante présente l’inertie expliquée de chaque axe factoriel (Série1) et la courbe du cumul des inerties ajustées des axes (Série2) :

Figure 2 : Histogramme des valeurs propres et courbe



Dans le même cadre, le tableau ci-dessus donne les valeurs propres associées à chaque axe ainsi que le taux d’inertie expliquée :

Tableau -3- : Les valeurs propres et pourcentage d’inertie des axes.

Axe	F1	F2	F3	F4	F5	F6	F7
Valeur Propre	0,158	0,121	0,113	0,107	0,103	0,088	0,086
Inertie (%)	7,351	5,601	5,241	4,953	4,795	4,107	4,008
% Cumulé	7,351	12,951	18,193	23,146	27,941	32,048	36,056
Axe	F8	F9	F10	F11	F12	F13	
Valeur Propre	0,084	0,081	0,080	0,078	0,077	0,077	
Inertie (%)	3,895	3,743	3,692	3,612	3,594	3,580	
% Cumulé	39,951	43,694	47,386	50,999	54,592	58,172	

En analysant la figure 2, on remarque une stationnarité de la courbe du cumul des inerties ajustées à partir du 12ème axe. De ce fait, et puisque l'apport des 13 premiers axes est significatif, nous les avons retenues parmi les 28 existants.

Surtout que le maximum d'inertie est expliqué par le premier axe et il ne dépasse pas 10% et les 13 premiers expliquent environ de 60% de l'inertie totale.

4- Résultats de l'AFD

L'analyse discriminante s'effectue après l'analyse des correspondances. En effet, nous avons appliqué l'AFD aux coordonnées des individus sur les 13 axes retenus dans l'analyse précédente. Les premiers résultats de cette analyse sont présentés dans le tableau de signification des axes suivant :

Tableau -4- : Signification des axes.

Axe	Lambda de Wilks	F	d.d.1 1	d.d.1 2	Signification
F1	0,778	24127,478	1	1	0,000
F2	0,571	31688,396	2	1	0,000
F3	0,482	30192,729	3	1	0,000
F4	0,449	25906,388	4	1	0,000
F5	0,428	22528,004	5	1	0,000
F6	0,416	19708,792	6	1	0,000
F7	0,405	17703,605	7	1	0,000
F8	0,395	16185,588	8	1	0,000
F9	0,386	14938,652	9	1	0,000
F10	0,380	13771,032	10	1	0,000
F11	0,379	12549,139	11	1	0,000
F12	0,379	11524,956	12	1	0,000
F13	0,379	10641,666	13	1	0,000

On remarque que le degré de signification de chaque axe est inférieur à 5% (égale à 0). Ainsi, les axes retenus dans la partie précédente sont tous discriminants. Ils s'agit des axes suivants : F1; F2; F3; F4; F5; F6; F7; F8; F9; F10; F11; F12; F13.

Le degré de discrimination des axes est aussi représenté par les scores qui varient entre -0,371 et 0,857 . Ces scores ainsi que les

degrés de signification nous ont permis d'éliminer les axes les moins discriminants, comme le nous l'avons déjà signalé. Nous avons obtenu le tableau suivant :

Tableau -5- : Score des axes.

Axe	Score
F1 3	- 0,049
F12	- 0,018
F11	- 0,043
F10	- 0,194
F9	0,292
F8	0,218
F7	- 0,223
F6	0,210
F5	0,155
F4	0,835
F3	0,857
F2	- 0,371
F1	0,586

A partir de ce dernier tableau, nous avons déterminé les scores des modalités et par la suite ceux des clients. En effet, et puisque les axes ne sont que des combinaisons linéaires des modalités, alors il devient facile de déterminer les scores des modalités à partir de ceux des axes. Ceci nous permettra de préciser les modalités les plus discriminantes pour chaque classe de variable.

5- Tableau des scores des modalités (Saporta, 2006)

Le score est une combinaison linéaire des coordonnées factorielles telles que les coefficients sont les scores des axes factoriels. Puisque les axes sont combinaison linéaire des modalités, alors le score s'exprime comme une combinaison linéaire des indicatrices des modalités, avec les coefficients sont les scores de ces modalités.

En effet, le score est obtenu à partir de la formule suivante :

$$\text{On a : } S = \sum_{j=1}^k d_j Z^j \quad \text{Avec : } Z^j = XU^j$$

Avec U^j sont les coordonnées des catégories sur le $j^{\text{ème}}$ axe.

$$\text{Alors, } S = \sum_{j=1}^k d_j Z^j = \sum_{j=1}^k d_j XU^j = X \underbrace{\sum_{j=1}^k d_j U^j}_{\text{grille de score}} \quad (1)$$

Avec :

- k : le nombre d'axes retenus ;
- X est le TDC :

$$X = \begin{matrix} & \begin{matrix} 1 & 0 & 0 & 0 & 0 & 1 \end{matrix} & \begin{matrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ n \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ & \cdot & & & \cdot & \\ & \cdot & & & \cdot & \\ & \cdot & & & \cdot & \\ & \cdot & & & \cdot & \end{pmatrix} & \begin{matrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{matrix} \end{matrix}$$

X : Variables indicatrices

Donc, d'après la formule (1), le score d'un individu est la somme des scores des modalités qui le caractérisent. Si on connaît les scores de toutes les modalités on peut déduire les scores des individus. Ainsi, en appliquant la formule (1), on trouve le tableau des scores des modalités présenté ci-dessous, il est appelé chez les professionnelles la grille de score. En effet, c'est le produit matriciel de la matrice constituée des coordonnées des modalités sur les axes significatifs et de la matrice constituée des scores de ces axes :

Tableau -6- : Score des modalités (Grille des scores).

Variable	Modalité	Score
ZONE	1	-0,0385
	2	-0,0718
	3	1,1935
	4	-0,4327
LOCAL	1	0,2098
	2	-0,3495
SEXE	1	0,0646
	2	-0,0718
AGE	1	0,1554
	2	-0,0365
	3	-0,1465
ETATMATR	1	0,0316
	2	-0,4500
	3	-0,0430
	4	-0,0531
SECTACT	1	-0,4402
	2	0,2565
	3	0,1918
	4	0,2904
TYPPRET	1	0,0246
	2	-0,2852

MNTACOR	1	-0,5070
	2	-0,1499
	3	0,0130
	4	0,2531
	5	-0,1860
	6	0,9722
PERIOD	1	0,7268
	2	1,0927
	3	-0,1279
DURPRET	1	1,0090
	2	-0,2993
	3	0,1521
TEMPATT	1	-0,5825
	2	-0,5460
	3	1,6072
DEGRATION	1	0,1722
	2	-0,0098
	3	-0,2597

D'après les tableaux 5 et 6, nous avons conclu que les modalités ayant des scores négatifs présentent un faible risque et donc caractérisent les «BONS », alors que les modalités ayant des scores positifs présentent un fort risque et donc caractérisent les « MAUVAIS ».

Ainsi, un client est jugé « BON » s'il a les particularités suivantes :

- ZONE : Nord-Est, Nord-Ouest, Sud-Ouest ;
- LOCAL : Rural ;
- SEXE : Femme ;
- $35 \leq \text{AGE} \leq 87$ ans ;
- ETATMATR : Divorcé, Veuf, Célibataire ;
- SECTACT : Agriculture ;
- TYPPRET : Prêt Individuel ;
- $1000 \leq \text{MNTACOR} < 4000 \text{ DH}$ ou $10\,000 \leq \text{MNTACOR} < 15\,000 \text{ DH}$;
- PERIOD : Mensuelle ;
- $13 \leq \text{DURPRET} < 19$ mois;
- TEMPATT d'une et deux semaines ;
- DEGRATIO > 0.

De ce fait, un « MAUVAIS » client se caractérise par le reste des modalités qui sont :

- ZONE : Sud-Est ;
- LOCAL : Urbain ;
- SEXE : Homme ;
- $18 \leq \text{AGE} < 35$ ans ;
- ETATMATR : Marié ;
- SECTACT : Commerce, Service, Artisanat ;
- TYPPRET : Prêt Solidaire ;
- $4000 \leq \text{MNTACOR} < 10\,000$ DH ou $15\,000 \leq \text{MNTACOR} < 50\,000$ DH ;
- PERIOD : Hebdomadaire ou Bimensuelle ;
- $5 \leq \text{DURPRET} < 13$ mois ou $19 \leq \text{DURPRET} < 61$ mois ;
- TEMPATT Entre deux et quatre semaines ;
- DEGRATIO = 0.

6- Résultat des tests de validation du modèle

(i) *Qualité d'affectation*

D'après le tableau ci dessous, nous remarquons que 92,15% des observations sont bien classées dans l'échantillon de base, et que le pourcentage des biens classés dans l'échantillon de base chez les «MAUVAIS» est de 80,67%, tandis qu'il est de 95,46% chez les «BON».

Nous concluons que le taux de bon classement et le taux d'erreur sont assez satisfaisants et permettent de dire que le modèle est acceptable.

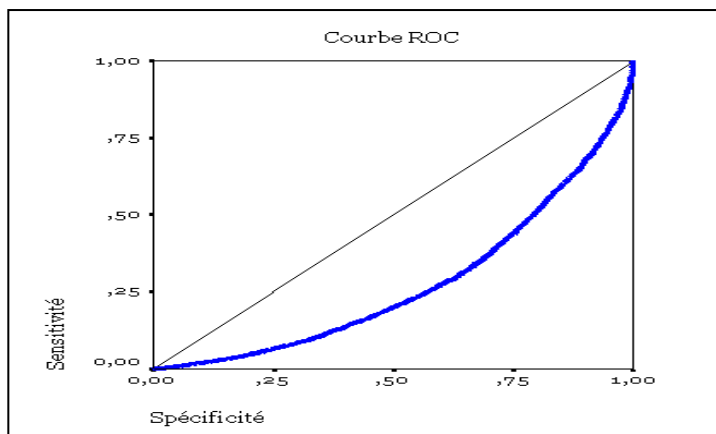
Tableau -7- : Les taux de classement des deux groupes de clients.

	Classes d'affectation prévues %		Total
	Défaillants	Bons	
Défaillants	80,67	19,33	100%
Bons	4,54	95,46	100%

(ii) *Courbe de ROC*

On constate dans la figure ci-dessous que la courbe de ROC s'éloigne suffisamment de la bissectrice, ce qui nous permet d'affirmer que le modèle est assez bon.

Figure -3- : Courbe de ROC du modèle disqual.



(iii) Indice de Gini

D'après les résultats figurant sur le tableau en dessous, nous calculerons l'indice de Gini (g):

On a : $g = (1 - 2 \cdot Z)$

avec Z est la zone sous la courbe, d'après les résultats affichés sur le tableau-8-, elle est égale à 0,276

Donc : $g = 44,8\%$.

La valeur de cet indice s'avère satisfaisante et confirme que le modèle est valide et que les résultats obtenus sont assez bons.

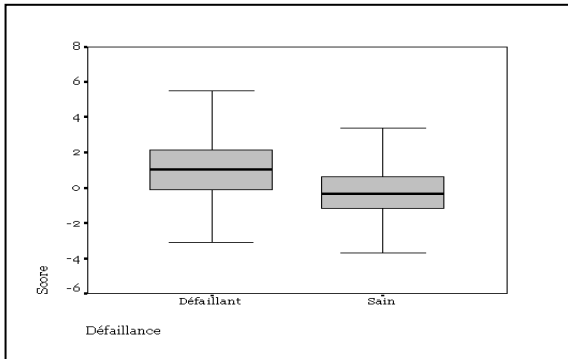
Tableau -8- : Calcul de la surface de la zone sous la courbe de ROC.

Zone	Erreur Std	Singnif. Asymptotique	Intervalle de confiance 95% asymptotique	
			Borne Inf	Borne Sup
0,276	0,002	0,000	0,272	0,280

(iv) Boîtes à moustaches

On constate, sur la figure-4-, que la boîte à moustaches représentant les « MAUVAIS » est assez écartée (vers le haut) de celle représentant les « BONS », cela nous permet de confirmer que nous avons une bonne discrimination entre les deux groupes.

Figure -4- : Boîtes à moustaches pour les deux groupes (cas du modèle disqual



7- Validation des scores par simulation

Après détermination du seuil d'affectation du clients suivant ses propres caractéristiques, cette validation est effectuée par simulation des nos propres affectation à celles constatées auprès de l'association Al amana.

(i) *Choix du seuil d'affectation des clients*

Nous avons choisi un seuil d'affectation des clients en se basant sur le taux de classement. En effet, nous avons choisi plusieurs seuils d'affectation en se basant sur les résultats des classes de score. Ensuite, nous avons calculé à chaque fois le taux de classement réalisé par ce seuil, le taux d'erreur, le taux de refus des «BONS» et le taux d'acceptation des «MAUVAIS» pour chacun de ces seuils. Alors, le seuil choisi est $S = 0,2$ possédant un taux de classement égale à 66,33% ; ainsi la prise de décision sera comme suit :

- Si le client a un score $\leq 0,2$ il est prédit «BON» ;
- Si le client a un score $> 0,2$ il est prédit «MAUVAIS».

(ii) *Simulation d'affectation par rapport à la base d'Al Amana***Tableau -9- :** Le nombre de clients bien classés par rapport au seuil $S=0,2$

			Défaillance		T O T A L
			Défaillants	Sains	
Seuil 0,2	Mauvais	Effectif	12 619	22 165	34 784
		% dans seuil 0,2	66,9	33,8	
	Bons	Effectif	6 246	43 361	49 607
		% dans seuil 0,2	33,1	66,2	
TOTAL		Effectif	18 865	65 526	84 391
		% dans seuil 0,2	100%	100%	

D'après ce tableau on calcule le taux de bon classement $(12619+43361)/84391=66,33\%$. Et le taux de différenciation $(6246+22165)/84391=33,67\%$. Ces taux donnent de bons résultats pour valider le seuil d'affectation choisi, mais ils s'avèrent insuffisants puisque ces résultats reposent uniquement sur l'échantillon utilisé pour l'estimation du modèle.

Il est donc suggéré à l'association de procéder à la mise à jour du seuil d'affectation en vérifiant à chaque fois le taux de bon classement.

Conclusion

Les caractéristiques des clients retenues dans notre analyse expliquent la séparation de la population d'étude en deux groupes : les « MAUVAIS » clients et les « BONS ».

D'autre part, les résultats du modèle de « scoring » par la méthode Disqual nous ont permis d'élaborer un système d'attribution de notes pour chaque client. A partir des caractéristiques de ce dernier, nous obtenons le score qui lui est associé en sommant les scores correspondants à ses modalités. L'affectation au groupe correspondant est ensuite effectuée en comparant le score obtenu avec le seuil d'affectation choisi. Ce dernier reste une décision stratégique.

Notons que le « scoring » est une technique qui nécessite un suivi régulier, afin d'actualiser les critères explicatifs, vérifier l'efficacité du score et l'exactitude des prévisions. Le chapitre suivant sera alloué à l'application d'une régression logistique comme autre technique pour le « Scoring ».

Références bibliographie

- Al Amana, Rapport d'activité 2006.
- Al Amana a dix ans, 2007.
- ALBAN THOMAS ; Économétrie des variables qualitatives ; DUNOD; Paris; 2000.
- Boubacar Diallo, Un Modèle de "Crédit Scoring" Pour une institution de micro-finance africaine: le cas de NYESIGISO au Mali, 2006.
- Churchill C. et Coster D. (2001) « CARE Manuel de gestion des risques en microfinance Indicateurs de Performance pour les Institutions de Microfinance », 2^e Edition
- Diallo B. (2006) « Un modèle de «Credit scoring pour une institution de microfinance Africaine, le cas de Nyesigiso au Mali »
- Fossou F. (2007) « Gestion stratégique et opérationnelle du Risque de Crédit dans les Institutions de Microfinance : cas du PADME-Benin »
- Gilbert SAPORTA; Analyse des données et statistiques ; DUNOD ; 1990.
- G. Saporta. Models for Understanding versus Models for Prediction . Inviting Conference COMPSTAT'08, Porto, Portugal, aug, pp. 315-322, Physica Verlag, 2008.
- Maddala G. S. (1996) : Limited Dependant and Qualitative Variables in Econometrics, Econometric Society Monograph, Cambridge UP , 401 p
- MIRIELLE BARDOS ; Analyse discriminante : application au risque et scoring financier, DUNOD, 2001.
- Muhammad Yunus and Karl Weber. Creating a World Without Poverty: Social Business and the Future of Capitalism, 2007.

- Schreiner Mark, le scoring peut-il attirer les investisseurs soucieux du profit vers le microcredit? 2007.
- Todd H. ed. ; Gibbons D.S. ; Pradesh A. ; Alam R. I. ; Khan S :Problems and Prospects of Group Lending in NGO Credit Programmes in India.;2001
- Todd, Women at the Center: Grameen Borrowers After One Decade. University Press Ltd., Dhaka, 1996.
- Yunus Muhammad, Banker to the Poor: Micro-Lending and the Battle Against World Poverty, 1999.
- Watson et Dunford, quels effets de l'IMF sur les conditions de vie ? Les analyses d'impacts, 2006
- Wonou.C. (2006) « Espace Micro finance N°002 : La gestion préventive et curative du risque de crédit dans les institutions de microfinance »

Sites Web

- <http://cedric.cnam.fr/~saporta/>.
- <http://cedric.cnam.fr/~saporta/discriminante.pdf>.
- <http://data.mining.free.fr>.
- <http://www.enterweb.org/microc-f.htm>.
- <http://www.planetfinance.org/>.

Annexes des données et des résultats

→ Annexes des Données :

- Zone géographique (zone) :

Tableau -1- : Codes de la variable zone géographique.

Zone	Nord -Est	Nord –	Sud -Est	Nord –
Code	1	2	3	4

Localisation (local) :

Tableau -2- : Codes de la variable localisation.

Localisation	Urbain	Rural
Code	1	2

genre (sexe) :

Tableau -3- : Codes de la variable genre.

Genre	Homme	Femme
Code	1	2

Âge (age) :

Tableau -4- : Codes de la variable age.

Age	[18-35]	[35-47]	[47-87]
Code	1	2	3

Etat matrimonial (etatmatr) :

Tableau -5- : Codes de la variable état matrimonial.

Etat	Marié	Divorcé	Veuf	Célibataire
Code	1	2	3	4

Secteur d'activité (sectact) :

Tableau -6- : Codes de la variable secteur d'activité.

Secteur	Agriculture	Commerce	Service	Artisanat
Code	1	2	3	4

- Type de prêt (typpret) :

- **Tableau -7- :** Codes de la variable type de prêt.

Tupe de prêt	Prêt solidaire	Prêt individuel
Code	1	2

- Montant accordé (mntacor) :

Tableau -8- : Codes de la variable des prêts accordé en milliers.

Prêt	[1-2[[2-4[[4-8[[8-10[[10-15[[15-50[
Code	1	2	3	4	5	6

- Périodicité de remboursement (period) :

Tableau -9- : Codes de la variable périodicité de remboursement

	hebdomadaire	Bimensuelle	Mensuelle
Code	1	2	3

- Durée de prêt en nombre de mois (durpret) :

Tableau -10- : Codes de la variable durée de prêt (en mois).

Durée de prêt	[5-13[[13-19[[19-61[
Code	1	2	3

- Temps d'attente en nombre de jours (tempatt) :

Tableau -11- : Codes de la variable temps d'attente.

	T<semaine	1s<T<2semaines	2s<T<4semaines
Code	1	2	3

- Degré de rationnement (degratio)

Tableau -12- : Codes de la variable degré de rationnement.

Degré de rationnement	0]0-1[>1
Code	1	2	3

- Défaillance (default) :

▪ **Tableau -13- : Codes de la variable défaillance.**

Défaillance	D	non D
Code	1	2

→Annexes des résultats

Tableau -1- : Résultats du test de Khi

variables	Khi-deux	Degré de liberté	Seuil critique
Zone	2678,95	3	0.000
Local	1153,73	1	0.000
Sexe	72,53	1	0.000
Age	311,76	2	0.000
Etatmatr	48,22	3	0.000
Sectact	1485,75	3	0.000
Typepret	6,009	1	0.014
Mntacor	454,033	5	0.000
Period	1656,77	2	0.000
Durpret	2330,37	1	0.000
Tempatt	4059,37	2	0.000
degratio	8,545	2	0.014

Tableau -2- : Résultats du test de Kruscal Wallis.

va	SKW	Degré de	Seuil
Zone	113,26	1	0.000
Local	1153,72	1	0.000
Sexe	72,53	1	0.000
Age	310,80	1	0.000
Etatmatr	36,38	1	0.000
Sectact	1183,14	1	0.000
Typepret	32,00	1	0.031
Mntacor	374,31	1	0.000
Period	1650,81	1	0.000
Durpret	2330	1	0.000
Tempatt	2970,27	1	0.000
degratio	27,43	1	0.041