
M.E.S., Numéro 128, mai - juin 2023

<https://www.mesrids.org>

Dépôt légal : MR 3.02103.57117

N°ISSN (en ligne) : 2790-3109

N°ISSN (impr.) : 2790-3095

Mise en ligne le 30 mai 2023



Revue Internationale des Dynamiques Sociales
Mouvements et Enjeux Sociaux
Kinshasa, mai - juin 2023

SEGMENTATION DES CANDIDATS SOLLICITANT DES PRETS DANS UNE IMF SELON LEURS DEGRES REELS DE CREDIBILITE RESPECTIFS

Apport de la Méthode K-means de Classification Automatique basée sur l'apprentissage non supervisé

par

Simon MAPHANA ma NGUMA

Professeur

Gabriel KHABI NDUDI

Divine MILUNDA MAKWENGE

Assistants

(Tous) *Faculté des Sciences Economiques et de Gestion*

Université de Kinshasa

Résumé

Cette étude propose à l'IMF ALPHA, ainsi surnommée pour des raisons d'anonymat et de respect de sa politique de confidentialité, une approche et un outil efficace pour fiabiliser davantage son processus d'octroi des prêts, jugé insuffisamment efficace.

Mots-clés : *microfinance, data mining, classification automatique, Intelligence Artificielle, apprentissage automatique, degrés de crédibilité.*

Abstract

This study offers the MFI ALPHA, thus nicknamed for reasons of anonymity and respect of its confidentiality policy, an approach and an effective tool to make its loan granting process more reliable, deemed insufficiently effective.

Keywords : *microfinance, data mining, automatic classification, Artificial Intelligence, automatic learning, degrees of credibility.*

INTRODUCTION

Dans son rapport annuel 2019 [65], la Banque Centrale du Congo, autorité de régulation du système financier en RDC et gestionnaire de la centrale des risques des Institutions de Microfinance (IMF), relève quelques faiblesses dans le processus d'octroi des prêts par les établissements de microcrédit, tout en avançant certains facteurs de cette situation. Celle-ci est, en effet, caractérisée par l'existence d'une proportion importante du portefeuille à risque (PAR), comprenant des créances douteuses qui totalisent plus de 30 jours de retard de paiement, soit une moyenne annuelle de 20,4 % entre 2015 et 2019, contre la norme maximale de 5,0 %. Bien que présentant une tendance générale à la baisse par la suite, tout en évoluant en dents de scie, ce taux est demeuré élevé, soit 9,5 % et 14,7 % respectivement en 2020 et 2021.

Ce constat alarmant corrobore les propos tenus par le Gouverneur de cette institution en reconnaissant publiquement, lors de l'ouverture de l'atelier organisé le 15 février 2017 à l'attention des représentants des IMF réunis à Kinshasa, que plusieurs établissements de microcrédit avaient vu leur rentabilité s'amenuiser sensiblement, entraînant même la faillite de certaines d'entre elles.

Les facteurs à la base de cette situation sont multiples : (i) l'absence de rigueur dans l'octroi des crédits, résultant d'une mauvaise appréciation de la solvabilité et de l'honorabilité des demandeurs des prêts, ainsi que dans le suivi des recouvrements des crédits consentis ; (ii) un interventionnisme abusif de certains responsables dans le processus d'octroi des crédits, aboutissant à des décisions prises parfois sur des bases peu objectives ; (iii) le manque de maîtrise, en particulier par les emprunteurs opérant dans le secteur quasi-informel, des outils de gestion efficaces susceptibles de garantir la viabilité

des affaires financées ; (iv) le non-respect, par certains débiteurs, de leurs engagements contractuels souscrits de payer régulièrement la quotité mensuelle due au titre de remboursement des dettes contractées auprès des IMF. Tous ces écueils, imputables à la fois aux responsables des IMF et aux emprunteurs, sont susceptibles de compromettre la rentabilité des activités financées et les chances de récupération des fonds prêtés.

Il ressort des considérations ci-dessus que la fiabilisation du processus d'octroi des crédits dans les IMF passe certes par une bonne gouvernance et une gestion orthodoxe de ces établissements, mais aussi par la détermination des degrés « réels » de crédibilité des clients sollicitant des prêts pour, d'une part, garantir la rentabilité des activités financées et, d'autre part, minimiser le risque de contrepartie, en n'accordant des prêts qu'à des candidats crédibles, c.à.d. solvables et respectueux des engagements souscrits vis-à-vis de leurs créanciers. Cette étude s'appesantit particulièrement sur la dernière préoccupation mise en exergue.

Par ailleurs, son champ d'investigation porte sur l'IMF ALPHA, ainsi surnommée pour des raisons d'anonymat et de respect de sa politique de confidentialité, mais qui a accepté de mettre à disposition les données ayant permis cette analyse. Elles sont extraites d'un échantillon représentatif de 100 dossiers de candidats ayant sollicité des prêts en 2022. On note tout de même que cette IMF passe pour la plus expérimentée et sert même de référence dans le secteur. En dépit de sa situation enviable, le constat alarmant fait ci-dessus par la BCC la concerne aussi : son taux de PAR se situe en effet dans la même fourchette que celui de l'ensemble du secteur et, par conséquent, au-dessus de la norme fixée à 5 %.

Dans ce contexte, la *question fondamentale à laquelle cette étude cherche à répondre* est la suivante : face aux écueils évoqués ci-dessus, comment l'IMF ALPHA peut-elle déterminer, avec le moins d'erreur possible, les degrés de crédibilité des candidats sollicitant des prêts, de manière à n'accorder des crédits qu'à ceux qui sont présumés solvables et respectueux de leurs engagements, à minimiser ainsi le risque relativement élevé de contrepartie qui caractérise son secteur d'activité, à garantir sa rentabilité et, par ricochet, sa pérennité ?

L'objectif global poursuivi dans cette étude est de proposer des stratégies à même rendre l'IMF ALPHA capable d'appréhender de manière pragmatique et objective les degrés de crédibilité des candidats emprunteurs, en proposant à cet effet une approche et des outils fiables d'aide à la décision permettant de répartir les candidats ayant des caractéristiques similaires dans des sous-groupes homogènes correspondant à leurs degrés réels de crédibilité.

De manière spécifique, l'étude doit :

- produire et mettre en compétition deux types de résultat de classification des candidats sollicitant des prêts auprès de l'IMF ALPHA : d'une part, celui issu de l'utilisation conjointe des approches empiriques ainsi que de la méthode statistique de « scoring » adoptées par cette IMF respectivement pour les nouveaux et les anciens clients, et, d'autre part, celui obtenu en recourant à la méthode K-means, un algorithme de classification automatique basé sur la technique d'apprentissage non supervisé, inspiré de l'Intelligence Artificielle ;
- apprécier leur fiabilité qui est tributaire de la qualité des approches et outils utilisés pour produire ces résultats ;
- choisir la meilleure alternative à recommander à cette IMF pour, d'une part, l'aider à ne consentir des prêts qu'à des candidats réellement crédibles, afin de minimiser le risque de contrepartie encouru, et d'autre part, déterminer les montants et délais de crédit à leur consentir par rapport à ceux sollicités.

La mise en œuvre correcte de cette alternative devrait, comme résultats attendus, rendre cette IMF à même :

- d'accélérer le processus de prise de décision en matière d'octroi des crédits ;
- de privilégier l'objectivité car la décision repose aussi bien sur les données réels en rapport avec les degrés de solvabilité des clients que sur leur honorabilité ainsi que celle de leurs parrains ;
- de réduire le risque de contrepartie en n'octroyant des prêts qu'à des candidats présumés solvables et jugés sérieux dans le respect de leurs engagements vis-à-vis des créanciers ;
- de mieux circonscrire, en conséquence, la hauteur des montants et les délais de remboursement des crédits à consentir.

En guise d'hypothèse de travail, on estime que les méthodes statistiques multivariées de « data mining » à but de classification automatique, basées sur l'apprentissage non supervisé relevant de l'Intelligence Artificielle, peuvent être mises à contribution pour aplanir les écueils évoqués ci-dessus et répondre ainsi efficacement à la question de la problématique. On pense en particulier à la méthode de partitionnement K-means. Ces méthodes sont d'autant mieux indiquées que la préoccupation analysée porte sur un problème multicritère qui, par définition, est complexe et donc difficile à résoudre. Cependant, au lieu d'utiliser à cet effet les méthodes mathématiques de recherche opérationnelle pour analyser les relations de préférence partielle, globale ou de sur-classement entre actions en procédant par des comparaisons binaires de ces actions, nous mettrons à contribution une méthode de statistique multivariée (K-means) pour réaliser la classification automatique. Dans ce contexte, les variables seront assimilées aux critères et les modalités des variables aux états des critères.

Le corps de ce texte est *structuré en quatre sections*, hormis l'Introduction : (i) cadre conceptuel de l'étude ; (ii) présentation de la méthode K-means ; (iii) présentation des informations et données recueillies ; (iv) interprétation et appréciation des résultats obtenus.

I. CADRE CONCEPTUEL DE L'ETUDE

Cette section s'ouvre par quelques précisions en rapport avec le sujet abordé. Il s'agit des précisions portant d'abord sur le champ d'investigation et le point focal de l'étude (1.1.), ensuite sur le caractère complexe de la problématique analysée (1.2.), enfin sur certains concepts permettant de mieux s'imprégner du processus de résolution efficace de cette problématique en mettant à contribution la méthode K-means.

1.1. Concepts en rapport avec le champ d'investigation et le point focal de l'étude

La microfinance est reconnue aujourd'hui comme une nouvelle niche de marché financier à travers laquelle la pauvreté pourrait être vaincue mondialement par un développement généralisé. Néanmoins le concept microfinance n'est pas aisé à définir à cause de sa complexité qui trouve son origine dans la diversité des formes et des pratiques qui caractérisent ce secteur de la science ; de sorte que sa définition ne fait guère l'unanimité des auteurs qui ont eu à s'y pencher. En résumé, et à la lumière de tous les doctrinaires référencés notamment dans [3], [13], [37], [64], [66] [67], [68], [85], [86], on peut penser que la quintessence du mécanisme de financement appelé « microfinance » peut, aujourd'hui, être aisément compris à travers les trois acceptions essentielles ci-après :

- dans son acception simple, il consiste en la mise en place des financements spécifiques (à savoir des microcrédits), en faveur des personnes « non-bancables » et pour des activités génératrices de revenus ;

- dans sa conception générale dans les pays en développement, il vise l'extension de la bancarisation des populations exclues du secteur bancaire, en leur offrant la possibilité de disposer d'un compte ;
- au-delà de la sécurisation de l'épargne et de l'octroi de crédits, les institutions de microfinance offrent de plus en plus des moyens de paiement, tels que des virements de fonds, nationaux ou internationaux.

Les IMF, point focal de l'étude, font partie du système financier d'un pays. Cependant, elles ont leurs particularités par rapport aux établissements bancaires classiques qu'on peut relever notamment dans les documents référencés dans [39], [52], [65] : crédit généralement à court terme, accordé avec la garantie des tiers ou avec caution solidaire. Par ailleurs, elles font face à un risque de contrepartie relativement plus élevé suite notamment au volume réduit, à la faible rentabilité, à la gestion parfois embryonnaire et peu orthodoxe des activités financées, ainsi qu'à l'honorabilité, parfois douteuse, des demandeurs des prêts, en particulier s'il s'agit de nouveaux candidats opérant de manière quasi-informelle. Elles ont donc besoin des méthodes aussi rigoureuses que celles de « data mining » pour évaluer automatiquement et efficacement le degré de crédibilité des demandeurs des prêts.

1.2. Précision sur la nature complexe du problème à résoudre

Le processus d'octroi des crédits établi par l'IMF ALPHA comporte plusieurs étapes : il débute en effet par la réception des demandes de crédit pour se terminer par la convocation des candidats emprunteurs en vue de la signature du contrat, en passant notamment par *l'analyse des dossiers par le Comité de crédit*. La préoccupation de cette étude tourne autour de cette étape intermédiaire à l'issue de laquelle deux séries de décisions doivent être prises : d'une part, rejet ou acceptation de la demande de prêt et, d'autre part, en cas d'acceptation, fixation du *montant* et du *délai* de crédit à consentir, par rapport à ceux sollicités.

Plusieurs aspects d'ordre administratif interviennent dans la prise de la première série des décisions et qui sont faciles à vérifier (documents administratifs requis, âge du demandeur des crédits, etc...). Mais en toile de fonds de la deuxième série des décisions, il y a l'épineux problème de la *détermination des degrés de crédibilité des candidats éligibles au crédit et, en conséquence, des niveaux de risque de contrepartie à supporter par l'IMF*. Ces paramètres constituent en effet la base objective de fixation des montants des crédits et des délais de remboursement des prêts accordés aux candidats emprunteurs, par rapport à ceux sollicités, afin de s'assurer du paiement régulier des acomptes mensuels attendus, de sauvegarder la rentabilité de l'établissements de microcrédit et, par ricochet, de garantir sa pérennité. La problématique sous-étude réside donc à ce niveau. A cet égard, la réponse à cette question est d'autant plus difficile à fournir que, comme dit dans l'Introduction, elle se pose dans le cadre d'un problème multicritère et, par conséquent, complexe.

En effet, parmi les nombreux facteurs contribuant à accroître le niveau de complexité des problèmes de gestion et de décision figure en bonne place le nombre de critères de décision. Les méthodes utilisées pour les résoudre efficacement conduisent à trois résultats possibles : (i) la constitution du noyau comprenant la ou les meilleures actions du lot, (ii) le classement de ces dernières en ordre utile ou (iii) leur regroupement dans des classes homogènes [38].

On peut résoudre un tel problème soit par les méthodes de recherche opérationnelle analysant les relations de préférence partielle, globale ou de sur-classement entre actions, en procédant à cet effet, par des comparaisons binaires de ces actions ; soit en mettant à contribution les méthodes statistiques multivariées de « data mining » pour réaliser la

classification automatique et/ou la prédiction. Dans cette étude, on exploitera l'une des méthodes relevant de cette deuxième approche, plus précisément l'algorithme de classification automatique K-means, utilisant la technique d'apprentissage non supervisé, inspirée de l'Intelligence Artificielle. Dans ce contexte, les variables seront assimilées aux critères et les modalités de ces variables aux états des critères.

1.3. Concepts pertinents en rapport avec la méthode K-means

1.3.1. Le « data mining » : définition et objectifs poursuivis

Le terme « data Mining » peut se traduire en français, selon les auteurs, par : forage des données, fouille des données ou exploration des données. D'autres auteurs parlent de « Knowledge Discovery in Data », ce terme pouvant être traduit littéralement en français par « découverte du savoir dans les données ». En fait le « data mining » représente l'ensemble des méthodes et techniques statistiques utilisées pour explorer de manière fiable de grandes masses des données caractérisées par plusieurs variables, afin d'en soutirer le maximum de renseignements utiles possibles. Ce concept est abondamment explicité dans plusieurs documents dont ceux référencés en [4], [7], [8], [16], [19], [23], [24], [26], [27], [31], [32], [33], [34], [35], [36], [51], [53], [58], [59], [62], [63], [69], [70], [71].

On peut considérer que l'activité de fouille des données poursuit quatre objectifs globaux : la description, la classification automatique, la régression et la prédiction. On s'appesantit ici uniquement sur l'objectif de classification automatique.

1.3.2. La classification automatique

Elle consiste à former dans un espace de représentation, des groupes homogènes des observations ayant des comportements similaires, en se basant sur les distances euclidiennes qui les séparent dans cet espace [82], [83]; ces groupes étant constitués uniquement au regard des valeurs des variables qui caractérisent ou décrivent ces observations. A cet égard, les hypothèses suivantes sont retenues : plus deux éléments sont proches dans l'espace multidimensionnelle, plus ils ont des chances de faire partie du même groupe ; par ailleurs, les groupes proches peuvent faire l'objet d'une agrégation pour former un seul groupe. Les groupes définitifs sont appelés, selon les auteurs, des classes, des familles, des catégories, des segments, si le nombre de groupes est prédéfini, ou des « clusters » dans le cas contraire.

De manière plus explicite, on peut définir la classification automatique comme suit.

Soient :

- $N = \{ e_1, e_2, \dots, e_i, e_j, \dots, e_n \}$: un ensemble fini d'éléments d'une population donnée ;
- $X = \{ 1, 2, \dots, j, \dots, p \}$: un ensemble fini de p variables pouvant caractériser ces éléments et constituant ainsi des critères de classification de ces éléments ;
- $P = \{ 1, 2, \dots, k, \dots, K \}$: une partition des N éléments composée de K groupes, tel que n_k est le nombre d'individus du groupe k .

Ainsi donc, classifier automatiquement N éléments d'une population donnée consiste, d'une part, à répartir ces éléments en K groupes disjoints selon p critères retenus, de manière à garantir une grande homogénéité au sein de chaque groupe et une hétérogénéité (différentiation ou discrimination) réelle entre groupes. A la fin du processus de classification, on attribue à chaque observation une étiquette identifiant son groupe d'appartenance.

1.3.3. L'Intelligence Artificielle

Il existe plusieurs publications traitant de l'Intelligence Artificielle (IA), référencées notamment en [1], [2], [5], [6], [9], [10], [15], [17], [25], [29], [42], [47], [48], [49], [50], [55], [60], [61], [72], [73], [87]. Les lignes qui suivent donnent quelques spécificités de ce concept, devenu aujourd'hui une véritable

discipline et un domaine de recherche très prisé, étant donné la vaste étendue de son champ d'application.

De manière plus simple, l'IA, basée sur le paradigme de l'intelligence humaine ou cervicale ou biologique, peut être définie comme la capacité dont on peut doter les ordinateurs (ou d'autres systèmes électroniques) de mimer, d'imiter ou de simuler l'intelligence de l'homme dans les tâches de perception des informations, de compréhension des messages, de raisonnement, de choix rationnel, de classification, de prédiction, etc..., faisant ainsi de ces machines de véritables outils d'aide à la résolution rapide des problèmes complexes et de prise de décision efficace, mieux que l'homme lui-même.

En tant que discipline ou domaine de recherche, l'IA est aussi définie comme l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence de l'homme. Alors que la première définition met l'accent sur l'*aptitude* dont jouit un ordinateur intelligent, la deuxième met plutôt en exergue les *techniques et technologies* qui concourent à rendre l'ordinateur intelligent. Comme on le constate, ces deux définitions se complètent harmonieusement.

1.3.4. L'apprentissage automatique

1.3.4.1. Définition du concept, objectifs poursuivis

Les spécificités de l'approche de classification ou de prédiction *basée sur l'apprentissage automatique* peuvent être trouvées notamment dans les documents référencés en [20], [28], [30], [40], [41], [43], [69], [73], [74], [75], [76], [77], [78], [79], [80], [81], [84].

Il en ressort, de manière générale, que l'apprentissage automatique ou apprentissage machine (« machine learning ») est une technique utilisée en Intelligence Artificielle (IA) pour rendre l'ordinateur capable :

- d'une part, d'apprendre par lui-même à connaître la structure d'un jeu des données quelconques soumises à son traitement, que ces dernières soient étiquetées ou pas ; et ce, grâce aux logiciels conçus à cet effet ; et,
- d'autre part, d'améliorer sa performance (c.à.d. à devenir intelligent), à l'issue de cet apprentissage (formation ou entraînement), pour :
 - soit classer automatiquement les individus ou objets présentant des caractéristiques similaires dans des groupes homogènes, tout en leur associant les étiquettes identifiant leurs groupes d'affectation respectifs ;
 - soit prédire, au regard des caractéristiques dont jouissent les observations d'un échantillon quelconque, leur classe d'appartenance, avec la probabilité associée ; et ce, à partir d'un autre jeu des données de référence déjà étiquetées selon la même logique que l'on voudrait appliquer pour estimer les valeurs des éléments de l'échantillon ou identifier leurs classes d'appartenance respectives.

On parle alors respectivement d'apprentissage non supervisé et supervisé. Ils sont utilisés pour réaliser respectivement la tâche de classification automatique et celle de prédiction. On ne s'appesantit spécifiquement ici que sur la technique d'apprentissage non supervisé qui sera mise à contribution dans le cadre de cette étude.

1.3.4.2. Apprentissage non supervisé

1.3.4.2.1. Principe général d'apprentissage non supervisé

Le recours à l'apprentissage automatique non supervisé suppose que les observations (individus ou objets) ne sont *ni classifiées d'avance, ni étiquetées en conséquence*. On utilise alors les algorithmes d'apprentissage automatique non supervisé pour analyser la structure de l'échantillon d'une population quelconque et constituer des groupes

homogènes de ses éléments ayant des caractéristiques similaires, tout en les étiquetant en conséquence afin de les distinguer (discriminer) facilement les uns par rapport aux autres.

La capacité de ces algorithmes à découvrir des similitudes et des différences dans les informations en fait la solution idéale pour notamment la segmentation des observations selon leurs caractéristiques.

1.3.4.2.2. Phases du processus d'apprentissage automatique non supervisé

Le processus utilisé par les algorithmes d'apprentissage automatique non supervisé pour entraîner l'ordinateur sur des ensembles des données non étiquetées et lui doter ainsi la capacité de les répartir dans des groupes homogènes comporte 4 phases :

Phase A : Constitution du jeu des données d'entrée non étiquetées, de taille N, en rapport avec le domaine sous étude et qui seront exploitées dans le cadre de l'analyse.

Phase B : Définition des règles de classification en rapport avec la problématique à résoudre (nombre de classes imposé ou pas).

Phase C : Choix de la méthode ou du modèle approprié, au regard de ses atouts.

Phase D : Entraînement de l'ordinateur sur des données non étiquetées et production, par l'algorithme ou modèle choisi, des résultats de la classification automatique, tout en étiquetant les observations selon leur classe d'affectation.

III. PRESENTATION DE LA METHODE K-MEANS

3.1. Quelques généralités sur les méthodes de classification automatique

On peut trouver les spécificités de ces méthodes dans les documents référencés en [11] , [12] , [14] , [21] , [22] , [31] , [32] , [33] , [36] , [44] , [45] , [46] , [56] , [82] , [83]. De manière générale, ces méthodes qui rentrent dans le champ de cette étude peuvent être regroupées en deux catégories : méthodes de classification automatique hiérarchique (ascendante et descendante) et méthodes de classification par partitionnement.

Avec les méthodes de la première catégorie, on suppose que les classes n'existent pas d'avance. Il faut alors les constituer à partir des algorithmes ne faisant intervenir que des données et non la subjectivité de l'expérimentateur. Par contre, celles de la deuxième catégorie partent d'un nombre de classes fixé au départ par l'expérimentateur, formant une partition des individus, et dans lesquelles on doit insérer au mieux les différents individus. Ces classes initiales peuvent être recomposées différemment au cours du processus pour obtenir une meilleure partition satisfaisant un critère donné. L'objectif est donc de trouver, parmi l'ensemble de toutes les partitions possibles, celle qui optimise un critère défini a priori, par exemple la minimisation de la distance entre individus. La méthode K-means utilisée ici fait partie de cette deuxième catégorie de méthodes puisqu'on considère que le nombre de classes est prédéfini dans l'IMF ALPHA.

3.2. Etapes de l'algorithme k-means ou k-moyennes [12], [31], [44], [82], [83]

Voici, de manière logique, les étapes de cet algorithme :

Etape 1 : Choix, opéré de manière aléatoire, de k individus comme centres initiaux des classes ;

Etape 2 : Attribution à chaque individu la classe la plus proche : ce qui définit au départ k classes ;

Etape 3 : Calcul des nouveaux centres d'inertie de chaque classe, les membres de chaque classe étant connus : on parle de recentrage ;

Etape 4 : Redistribution des individus dans la classe qui leur est la plus proche en tenant compte des nouveaux centres de classe calculés à l'étape précédente ;

Etape 5 : Retour à l'étape 3 jusqu'à ce qu'il y ait convergence, c'est-à-dire jusqu'à ce qu'il n'y ait plus aucune possibilité pour un individu de changer de classe.

3.3. Modalités de présentation des résultats

Les résultats peuvent être présentés sous deux formes : un tableau reprenant à côté de chaque individu sa classe d'appartenance ; ou bien un diagramme de visualisation de la répartition des individus entre les k classes définies par l'expérimentateur, tel que illustré en Annexe 1.

IV. INFORMATIONS ET DONNEES RECUEILLIES AUPRES DE L'IMF ALPHA

I.1. Approches de catégorisation des candidats emprunteurs

Ces approches diffèrent selon que les candidats sollicitant des prêts sont des anciens ou des nouveaux clients.

I.1.1. Cas des anciens clients

L'IMF ALPHA utilise le logiciel ARGOS, un logiciel de « credit scoring », pour classer les anciens clients. Son fonctionnement est décrit ci-après :

- les données caractérisant les anciens clients sont importées dans le logiciel : il s'agit des montants des crédits accordés, des dates de remboursement et du nombre de jours de retard pour chaque remboursement ;
- sur base de ces données, le logiciel génère le score comportemental afférent au respect des échéances et des quotités mensuelles payées ;
- s'appuyant sur ce score, le logiciel classe les anciens clients en 3 catégories selon le degré de risque qu'ils présentent : « High Risk » (C1), « Low-Mid Risk » (C2), « Low Risk » (C3) ;
- le paramétrage du logiciel est tel qu'il limite le montant minimum et maximum à allouer respectivement à 500 \$ et 5.000 \$, pour une durée de 11 mois.

Il y a lieu de noter que le score lié à chaque ancien client permet en même temps de décider du renouvellement ou pas du crédit ainsi que, en cas d'avis favorable, du montant à accorder.

En ce qui concerne les nouveaux clients, le logiciel se limite à recueillir, puis à reproduire fidèlement la situation des ratios de solvabilité des candidats, élaborés à partir des données financières en rapport avec leurs activités. Les analystes devront alors exploiter ces données pour décider de la classe d'appartenance de ces candidats.

I.1.2. Cas des nouveaux clients

Selon les explications obtenues des personnes impliquées dans la décision d'octroi des prêts, deux principales exigences président à la détermination des degrés de crédibilité des candidats emprunteurs : leur capacité de remboursement et leur honorabilité dans le paiement des quotités mensuelles dues. La capacité est appréciée sur base des valeurs des ratios de solvabilité, calculées à partir des données financières fournies par les candidats ; et ce, en se référant, pour leur exploitation, aux autres approches empiriques suivantes : l'allure générale des valeurs des ratios reflétant le degré de solvabilité par critère, les valeurs des ratios observées au niveau du (des) critère(s) jugé(s) les plus importants, ainsi que la moyenne des valeurs des ratios, calculée, pour chaque candidat, sur l'ensemble des critères considérés. L'honorabilité des candidats par contre s'apprécie sur base du respect des engagements souscrits avec leurs partenaires commerciaux, financiers ou même vis-à-vis du public en général. Tenant compte de tous ces paramètres, les analystes peuvent classer les nouveaux candidats dans C1, C2 ou C3 selon qu'ils sont considérés comme douteux, peu crédibles ou très crédibles, c.à.d. selon qu'ils présentent un niveau de risque élevé, moyen ou faible.

I.2. Approches de détermination des montants et délais de crédit à consentir

La détermination du montant du prêt et du délai de paiement à consentir par rapport à ceux sollicités se fait de manière empirique, en appréciant le degré d'adéquation entre, d'une part, le montant et le délai sollicité et, d'autre part, le degré de crédibilité reconnue au nouveau client qui reflète la capacité réelle de remboursement de sa dette. De manière générale, les principes qui guident les membres du Comité d'octroi des prêts pour décider des valeurs de ces deux variables de décision sont décrits ci-après pour :

- les candidats « douteux » (catégorie C1) : réduire à la fois le montant et l'échéance sollicités de manière à parvenir à une baisse sensible de la quotité mensuelle due.
- les candidats « peu crédibles » (catégorie C2) : réduire uniquement le montant sollicité et garder inchangé le délai de remboursement sollicité, entraînant automatiquement quelque peu, de ce fait, la diminution de la quotité mensuelle due.
- les candidats « crédibles et très crédibles » (catégorie C3) : trois alternatives sont envisageables pour cette catégorie de clients dont la capacité de paiement des acomptes mensuels est jugée élevée, voire très élevée :
 - (i) accorder l'intégralité du montant et du délai de crédit sollicités, avec pour effet le maintien de la quotité mensuelle due ;
 - (ii) revoir à la baisse uniquement le délai de remboursement sollicité, tout en gardant inchangé le montant du prêt sollicité ; cet exercice devant aboutir à la hausse de la quotité mensuelle due ;
 - (iii) de manière exceptionnelle, accroître le montant accordé, tout en maintenant ou en réduisant le délai de paiement consenti, entraînant ainsi automatiquement une augmentation très sensible de la quotité mensuelle due.

I.3. Critères d'appréciation des candidats

Les dix critères d'appréciation de la solvabilité des demandeurs individuels des prêts ainsi que quelques indications sur leurs spécificités sont repris dans le tableau ci-dessous.

Tableau I. Présentation des critères d'appréciation des dossiers des demandeurs des crédits auprès de l'IMF ALPHA

N°	Libellé	Formule (ratio ou rapport)	Norme	Observation
1	Ratio du Taux d'endettement	$R_1 = \text{Total des dettes (y compris l'emprunt proposé)} / \text{Capital Propre}$	< 50 %	<i>d.d.</i>
2	Ratio du Service de la dette assuré par le résultat réalisé sur l'activité financée par l'IMF ALPHA	$R_2 = \text{Acompte mensuel en remboursement du crédit} / \text{Profit net mensuel}$	< 50 %	<i>d.n.d.</i>
3	Ratio du Service de la dette assuré par le résultat réalisé sur l'ensemble des affaires du client	$R_3 = \text{Acompte mensuel en remboursement du crédit} / \text{Capacité mensuelle de paiement}$	< 70 %	<i>d.d.</i>
4	Nombre de jours ouvrables pour collecter le montant de l'acompte mensuel	$R_4 = (\text{Acompte mensuel} / \text{Chiffre d'affaires}) \times \text{nombre de jours ouvrables du mois}$	< 7 jours	<i>d.n.d.</i>
5	Prêt sur actifs circulants	$R_5 = \text{Prêts} / \text{Total Actifs Circulants}$	≤ 1	<i>d.d.</i>
6	Ratio de Liquidité	$R_6 = \text{Actifs Circulants} / (\text{Dettes à court terme} + 3 \text{ prochaines mensualités du prêt})$	≥ 1	<i>d.d.</i>
7	Ratio de couverture du prêt par la garantie offerte	$R_7 = \text{Somme des gages évalués au prix du marché} / \text{Prêt IMF}$	≥ 80 %	<i>d.d.</i>
8	Ratio de Marge de profit net	$R_8 = \text{Profit Net} / \text{Chiffre d'affaires}$	<i>d.n.d.</i>	<i>d.n.d.</i>
9	Ratio de Marge de profit brut	$R_9 = \text{Profit Brut} / \text{Chiffre d'affaires}$	≥ 20 %	<i>d.d.</i>
10	Vitesse de Rotation des stocks	$R_{10} = \text{Coût des Marchandises Vendues par mois} / \text{Actif circulant}$ $R_{10} = \text{Stock vendu par mois} / \text{Stock moyen durant le mois}$	> 1	<i>d.d.</i>

Source : Enquête auprès des différentes IMF de Kinshasa

Seuls les sept critères pour lesquels les données sur le degré de solvabilité de l'ensemble des candidats de l'échantillon sont disponibles (*d.d.*) seront considérés ici. Par ailleurs, on doit aussi tenir compte d'un 8^{ème} critère relatif à l'honorabilité des clients. L'indice R11 se réfère donc à ce dernier critère ; sachant que l'honorabilité de la personne présentée comme caution morale peut se répercuter sur celle du demandeur parrainé.

I.4. Echantillon des candidats emprunteurs et leurs caractéristiques

Les données chiffrées fournies par l'IMF ALPHA ont été extraites des 100 dossiers des demandeurs de crédit, cette IMF n'ayant pas souhaité aller au-delà de ce nombre compte tenu du caractère sensible de ce type d'informations.

La taille de la population cible (nombre total des dossiers) n'étant pas connu, le recours à la méthode de sondage aléatoire était exclu. Les dossiers ont donc été sélectionnés de manière empirique ou raisonnée, en utilisant toutefois une variante de « la technique des itinéraires » consistant à retenir dans l'échantillon le dossier du x^{ème} candidat figurant sur la liste générale des candidats, « x » étant le pas choisi. La technique des itinéraires donne des résultats satisfaisants dans la mesure où elle s'apparente au sondage ou tirage aléatoire systématique, bien que la taille de la population de base ne soit pas connue d'avance.

On a ainsi pu retrouver dans l'échantillon retenu des clients concernés par les 4 types de décision que le Comité d'évaluation est appelé à prendre concernant le montant et le délai de crédit à consentir par rapport à ceux sollicités, selon le niveau de crédibilité présumé du candidat, à savoir : maintien du montant et du délai ; maintien du montant et réduction du délai ; réduction du montant et maintien du délai ; réduction du montant et du délai.

Enfin, la taille de l'échantillon retenue (100 dossiers) peut paraître relativement réduite au regard de l'exigence de « big data » que requiert l'usage de la technique d'apprentissage automatique. De tels jeux de données ne sont pas toujours faciles à constituer dans la pratique. C'est pourquoi, tout en reconnaissant que les grandes masses des données contribuent à améliorer la précision des modèles d'apprentissage automatique, on note que les organisations n'ont pas toujours besoin de se doter des très grandes quantités des données pour utiliser les techniques d'apprentissage automatique. Le secret réside tout simplement dans la coordination : les « data miners » doivent utiliser les données les plus appropriées (c.à.d. pertinentes, précises et représentatives des caractéristiques dont jouissent les éléments de l'échantillon retenu), les bons algorithmes, les bons modèles et les bons logiciels d'exploitation des données ^[80].

I.5. Evaluations des candidats

Les évaluations des demandeurs des crédits se présentent sous forme de ratios ou rapports traduisant leurs degrés de solvabilité ainsi que sous forme d'indice de leurs degrés d'honorabilité respectifs. Les ratios ont été calculés en tenant compte, non seulement des informations tirées des dossiers des candidats, mais aussi des données récoltées sur terrain par les analystes de l'IMF ALPHA. (voir Tableau en Annexe 1). Ce tableau renseigne également sur la classe d'appartenance de chaque candidat ainsi que les montants sollicités et accordés, en dollars, les délais de remboursement sollicités et accordés, en mois, en application des approches définies respectivement aux § 3.1.1 et § 3.1.2.

I.6. Mise des données recueillies sous une forme exploitable

Les ratios calculés sur base des formules définies pour chaque critère étant des valeurs continues, on devra les « discrétiser » pour mieux faire ressortir les différents degrés de crédibilité, étant donné la présence, sur la liste, à la fois des critères maximisant

et minimisant. A cet égard, on opte pour un découpage binaire de ces séries des valeurs continues en procédant comme suit, tel que préconisé dans le document référencé en [42].

1. Pour les critères (variables) maximisant

- calcul des valeurs de coupure choisies :

V1 = valeur minimale de la série, correspondant à la norme

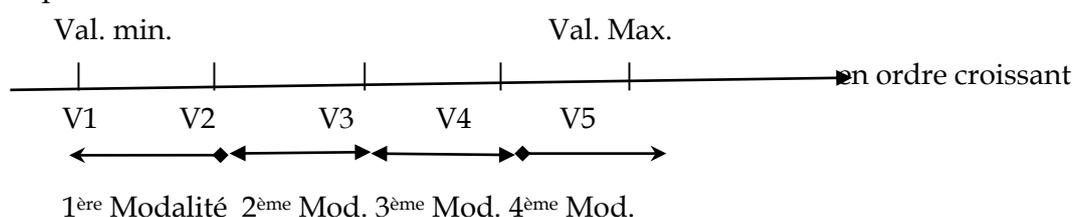
V5 = valeur maximale observée de la série

$$V3 = (V1 + V5) / 2$$

$$V2 = (V1 + V3) / 2$$

$$V4 = (V3 + V5) / 2$$

Schématiquement, on a :



- Conditions d'appartenance à chaque modalité :

Soit X représentant une valeur de la série. Les conditions d'appartenance aux différentes modalités seront définies comme suit :

Modalité 1 : $X < V2$

Modalité 2 : $V2 \leq X < V3$

Modalité 3 : $V3 \leq X < V4$

Modalité 4 : $X \geq V4$

2. Pour les critères (variables) minimisant

- calcul des valeurs de coupure choisies :

V1 = valeur minimale de la série

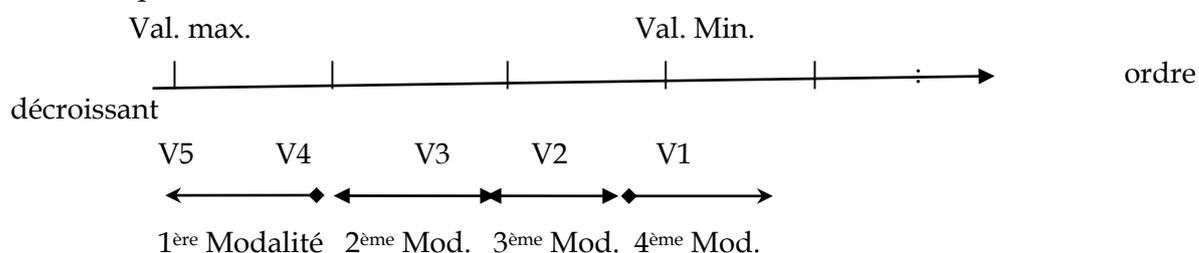
V5 = valeur maximale de la série, correspondant à la norme

$$V3 = (V1 + V5) / 2$$

$$V4 = (V3 + V5) / 2$$

$$V2 = (V1 + V3) / 2$$

Schématiquement, on a :



- Conditions d'appartenance à chaque modalité :

Modalité 1 : $X > V4$

Modalité 2 : $V3 < X \leq V4$

Modalité 3 : $V2 < X \leq V3$

Modalité 4 : $X \leq V2$

Ainsi donc, pour une variable donnée, si la valeur du ratio s'éloigne de la norme dans le sens favorable, le numéro d'ordre de la modalité correspondant et, par ricochet, le degré de solvabilité du client à évaluer augmente aussi. Le Tableau 2 définit concrètement les modalités de chaque variable :

Tableau II. Détermination des modalités des variables numériques discrétisées

Variables	Norme	Nature	Valeurs de coupure	Modalités (Intervalles)	N° d'ordre des modalités (par préférence croissante)
R ₁	< 50 %	Minimisant	V5 = 0,50 V4 = 0,38 V3 = 0,25 V2 = 0,13 V1 = 0,00	X > 0,38	1
				0,25 < X ≤ 0,38	2
				0,13 < X ≤ 0,25	3
				X ≤ 0,13 %	4
R ₃	< 0,70	Minimisant	V5 = 0,70 V4 = 0,53 V3 = 0,36 V2 = 0,19 V1 = 0,01	X > 0,53	1
				0,36 < X ≤ 0,53	2
				0,19 < X ≤ 0,36	3
				X ≤ 0,19	4
R ₅	≤ 1,00	Minimisant	V5 = 1,00 V4 = 0,82 V3 = 0,63 V2 = 0,44 V1 = 0,25	X > 0,82	1
				0,63 < X ≤ 0,82	2
				0,44 < X ≤ 0,63	3
				X ≤ 0,44	4
R ₆	≥ 1,00	Maximisant	V1 = 1,00 V2 = 3,90 V3 = 6,80 V4 = 9,70 V5 = 12,60	X < 3,90	1
				3,90 ≤ X < 6,80	2
				6,80 ≤ X < 9,70	3
				X ≥ 9,70	4
R ₇	≥ 0,80	Maximisant	V1 = 0,80 V2 = 0,98 V3 = 1,15 V4 = 1,33 V5 = 1,50	X < 0,98	1
				0,98 ≤ X < 1,15	2
				1,15 ≤ X < 1,33	3
				X ≥ 1,33	4
R ₉	≥ 20 %	Maximisant	V1 = 0,02 V2 = 0,24 V3 = 0,45 V4 = 0,67 V5 = 0,88	X < 0,24	1
				0,24 ≤ X < 0,45	2
				0,45 ≤ X < 0,67	3
				X ≥ 0,67	4
R ₁₀	> 1,00	Maximisant	V1 = 0,26 V2 = 1,00 V3 = 1,75 V4 = 2,50 V5 = 3,24	X < 1,00	1
				1,00 ≤ X < 1,75	2
				1,75 ≤ X < 2,55	3
				X ≥ 2,50	4
					T

Source : élaboré sur base des données du Tableau en Annexe 2.

Par ailleurs les modalités retenues pour le critère R11 sont 1, 2 et 4 correspondant respectivement à l'état « pas honorable », « honorable » et « très honorable ». Ces données seront donc codifiées suivant les numéros d'ordre des modalités définies dans le Tableau

ci-dessus. La base des données d'input au modèle de segmentation K-means est reprise en Annexe 3.

V. PRESENTATION ET APPRECIATION DES RESULTATS OBTENUS

On présente ici, tout en les appréciant, le résultat de classification issu de la mise en œuvre des approches en vigueur au sein de l'IMF ALPHA et celui produit par l'algorithme K-means.

5.1. Classification fournie par les approches empiriques en vigueur au sein de l'IMF ALPHA

5.1.1. Résultat de la classification obtenu

La répartition des 100 candidats demandeurs des prêts dans les 3 groupes retenus, conformément aux approches et principes décrits respectivement au § 3.1.2. et au § 3.2., est renseignée dans l'Annexe 2. Il en ressort la synthèse suivante.

Tableau III. Synthèse de la répartition des clients selon la classification empirique de l'IMF ALPHA

Catégories de clients	Effectif	Pourcentage
clients douteux	44	44 %
clients peu crédibles	31	31 %
clients très crédibles	25	25 %
Total	100	100 %

Source : Tableau en annexe 1 ou 2

La majorité des candidats (44 %) sont donc considérés comme douteux, 31 % peu crédibles et 25 % très crédibles.

5.1.2. Appréciation du résultat de la classification obtenue selon les approches utilisées par l'IMF ALPHA

5.1.2.1. Appréciation du résultat de la classification des anciens clients

Ce résultat est fourni par le logiciel de « credit scoring » ARGOS. Le recours à ce logiciel présente certes quelques avantages mais aussi et surtout plusieurs inconvénients.

On peut retenir les deux avantages suivants : rapidité dans le traitement des renouvellements des crédits, l'intervention du Comité d'analyse des dossiers n'étant pas requise ; objectivité dans le renouvellement des crédits et dans la détermination des montants à accorder. En ce qui concerne les inconvénients, on note ce qui suit : utilisation du logiciel limitée aux seuls anciens clients ; montant de crédit accordé compris dans l'intervalle de [500\$; 5.000\$], alors que le candidat peut souhaiter aller en deçà ou au-delà des limites minimale ou maximale fixées ; détermination des degrés de crédibilité des anciens clients basée uniquement sur le comportement passé de ces derniers quant au respect de leurs engagements contractuels ; une telle vision des choses suppose que le comportement, aussi flatteur soit-il, affiché antérieurement par cette catégorie de clients justifierait ou déterminerait aussi automatiquement celui qu'ils adopteraient ultérieurement : ce qui n'est pas toujours évident.

5.1.2.2. Appréciation du résultat de la classification des nouveaux candidats

Cette classification résulte de la mise en œuvre des approches empiriques évoquées au § 3.1.2. Le recours à ces approches comporte plusieurs faiblesses.

- *Concernant l'approche basée sur l'allure générale des valeurs des ratios de solvabilité ainsi que celle recourant, pour chaque candidat, à la valeur moyenne des ratios calculée sur l'ensemble des critères : Possibilité d'injecter une certaine dose de subjectivité dans le processus de*

prise de décision ; risque de prendre des décisions biaisées du fait de la présence concomitante des critères maximisant et minimisant sur la liste des points de vue retenus : en effet, une même valeur très élevée (ou très faible) de ratio portant sur deux critères différents ne peut être interprétée de la même manière si l'un des critères est maximisant et l'autre minimisant.

- *Concernant l'approche basée sur les critères jugés les plus importants :*

Prise en compte d'un seul ou de quelques critères, excluant d'office tous les autres du processus de prise de décision.

Outre les inconvénients évoqués ci-dessus, il s'ajoute un autre argument mettant en exergue les faiblesses de la pratique empirique en vigueur au sein de l'IMF ALPHA en matière d'octroi des prêts. Il s'agit du taux moyen de son PAR qui se situe dans la même fourchette que celui de l'ensemble du secteur et, par conséquent, au-dessus de la norme fixée à 5 %. Pour la BCC, un tel niveau du taux du PAR résulte aussi d'une absence de rigueur dans l'appréciation des degrés réels de crédibilité des candidats emprunteurs.

En conclusion, ces différents constats portent à croire que certaines décisions d'octroi des prêts au sein de l'IMF ALPHA ne sont pas prises avec toute la rigueur et l'objectivité souhaitées ; de sorte que le résultat de classification qui en résulte devrait être considéré comme sujet à caution.

5.2. CLASSIFICATION FOURNIE PAR LA METHODE K-MEANS

5.2.1. Résultat de la classification obtenue

L'algorithme K-means, exploité par le logiciel « Orange Data » [18], fournit les résultats de la classification automatique aussi bien sous forme graphique que dans un tableau à double entrée, comme l'indique le schéma suivant, en mettant à contribution la technique d'apprentissage automatique non supervisé.

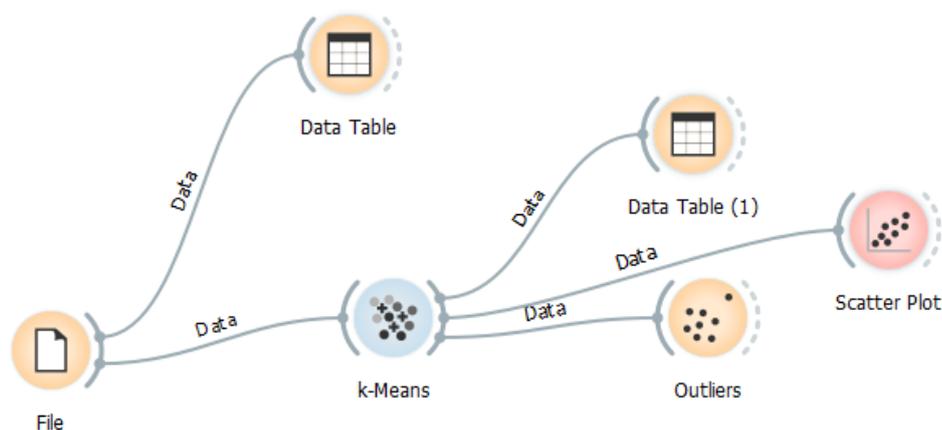


Fig. 3. : schéma logique d'obtention des résultats de la classification par K-means, d'après « Orange data »

où :

- File : dataset (fichier de base) importé dans son emplacement dans la machine;
- Data Table : Présentation des données dans une table avant la segmentation ou des résultats après la segmentation ;
- Scatter Plot : Présentation des résultats sous forme graphique après la segmentation ;

- Outliers : présentation des données ayant un comportement atypique par rapport aux autres : dans cette étude, la liste des données atypiques (outliers) est vide dans le cas sous étude.

La visualisation graphique de la classification obtenue dans le diagramme suivant met en exergue les 3 classes d'affectation des candidats emprunteurs.

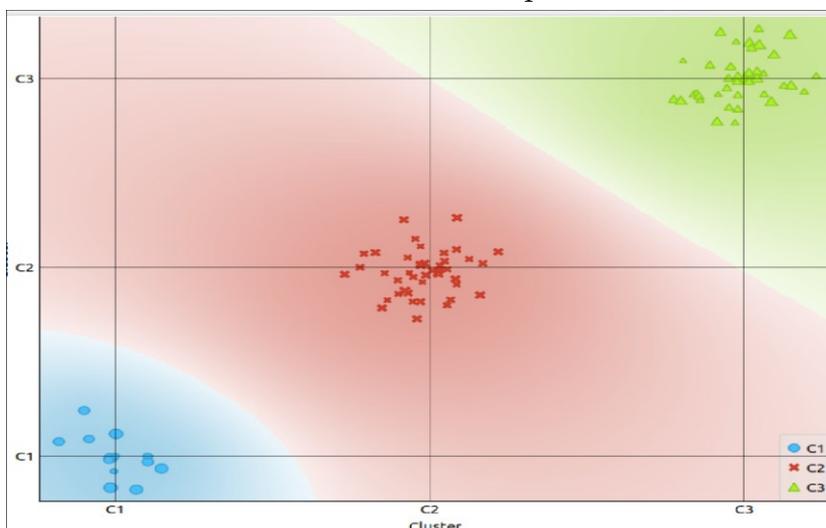


Fig 4. Représentation graphique de la classification fournie par Orange Data exploitant la méthode K-Means

Les résultats détaillés de la classification produite par la méthode K-means sont présentés dans le tableau en Annexe 3, à côté de ceux fournis empiriquement par l'IMF ALPHA, afin de faciliter la comparaison entre les deux approches de segmentation des candidats. En voici cependant la synthèse dans le tableau suivant.

Tableau IV. Synthèse de la répartition des clients selon la la méthode de classification automatique K-means

Catégories de clients	Effectif	Pourcentage
clients douteux	13	13 %
clients peu crédibles	43	43 %
clients très crédibles	44	44 %
Total	100	100 %

Source : Tableau en annexe 2

La majorité des candidats (44 %) sont donc considérés comme très crédibles (C3), suivis de très près par ceux présumés peu crédibles (43 %) (C2), puis, loin derrière, par ceux jugés douteux qui ne représentent que 13 % (C1).

5.2.2. Appréciation du résultat de la classification fournie par K-means

Le résultat de la classification produit par la méthode K-means présente certains atouts qui plaident en sa faveur.

- d'abord, cette méthode n'exploite que des données d'input tirées des dossiers des candidats, et qui sont censées refléter leurs degrés réels de solvabilité et d'honorabilité. En principe, elle met l'IMF à l'abri de toutes considérations subjectives.
- ensuite, les données en valeurs continues devant être exploitées par cette méthode ont préalablement fait l'objet d'une discrétisation différenciée, selon qu'elles se rapportent à

un critère maximisant ou minimisant, pour éviter tout biais dans la détermination et l'interprétation des degrés de crédibilité.

- enfin, l'Annexe 3 met en exergue le cas de quelques clients qui, au regard de l'allure des évaluations discrétisées renseignées pour les différents critères, peuvent logiquement être considérés comme crédibles ou très crédibles (C3), ainsi que le prévoit K-means, mais dont la classification en C1 (clients douteux) selon l'IMF ALPHA ne semble logiquement pas se justifier (voir les clients marqués par un astérisque à la dernière colonne).

5.3. Recommandation finale

Les deux résultats de classification des candidats présentés respectivement dans les deux tableaux synthèses 3 et 4 sont presque diamétralement opposés. En effet, pour l'IMF ALPHA, 44 % des clients sont douteux, 31 % peu crédibles et 25 % très crédibles, contre respectivement 13 %, 43 % et 44 % selon la méthode K-means. Cette situation est confirmée par le faible taux de classification ex-aequo, soit 30 %, (calculé à partir des effectifs situés sur la diagonale principale de la matrice de confusion du Tableau 5, contre un taux d'erreur ou de classifications divergentes élevé de 70 %, obtenu par différence.

Tableau V. Tableau croisé entre classifications selon l'IMF ALPHA et la méthode K-means

Résultats des classifications						
	Effectif	Types	Classification empirique de l'IMF ALPHA			Total
			C1. Douteux	C2. Peu crédibles	C3. Très crédibles	
Classification Automatique Par la méthode de K-means	Effectif	C1. Douteux	<u>11</u>	1	1	13
		C2. Peu crédibles	10	<u>14</u>	19	43
		C3. Très crédibles	23	16	<u>5</u>	44
		Total	44	31	25	100

Source : Annexe 3.

Au regard de ces résultats contrariés, il se pose un problème de choix entre celui issu de la pratique en vigueur au sein de l'IMF ALPHA et celui produit par la méthode K-means.

Sur base des appréciations faites aux § 5.1.2. et 5.2.2., on peut logiquement recommander à l'IMF ALPHA de privilégier le résultat de la classification fourni par la méthode K-means par rapport à celui issu de la pratique en vigueur au sein de cette IMF. Il découle des analyses ci-dessus qu'un processus fiabilisé de détermination des degrés de crédibilité de candidats sollicitant des prêts devrait se dérouler selon les étapes suivantes :

- détermination des niveaux de risque (au moins deux) auxquels l'IMF s'attend à faire face, tout en les spécifiant (exemple : élevé, moyen, faible) ; ces derniers correspondent, mutatis mutandis, aux différents degrés de crédibilité attendus des candidats sollicitant des prêts (exemple : douteux, peu crédible, très crédible) et donc au nombre de classes d'affectation possibles de ces candidats ;
- définition, dans la foulée, des principes de fixation des montants et délais de crédit à consentir par rapport à ceux sollicités, selon la classe d'appartenance des candidats emprunteurs ;
- choix des critères d'appréciation de la crédibilité des candidats sollicitant des prêts ;

- constitution de l'échantillon des candidats emprunteurs à classer, avec leurs caractéristiques (évaluations sous forme de ratios reflétant leurs degrés réels de solvabilité par critère ainsi que leur honorabilité) ;
- discrétisation, de manière différenciée, des valeurs continues (ratios de solvabilité) représentant les évaluations des candidats, selon qu'elles se rapportent aux critères maximisant ou minimisant ;
- mise à contribution, par le Comité d'analyse de dossiers, d'une méthode de segmentation automatique par partitionnement, utilisant la technique d'apprentissage non supervisé, à l'instar de la méthode K-means, pour classer de manière fiable les candidats emprunteurs de l'échantillon, *anciens comme nouveaux*, dans leurs groupes d'appartenance, conformément à leurs degrés de crédibilité respectifs ;
- détermination, sur cette base, des montants et délais de crédit à leur accorder par rapport à ceux sollicités, selon les principes définis au § 3.2.

CONCLUSION

Pour rappel, et de manière explicite, cette étude s'est assigné un triple objectif :

- produire et mettre en compétition deux types de résultat de classification des candidats sollicitant des prêts auprès de l'IMF ALPHA : d'une part, celui issu de l'utilisation conjointe des approches empiriques ainsi que de la méthode statistique de « scoring » adoptées par cette IMF respectivement pour les nouveaux et les anciens clients, et, d'autre part, celui obtenu en recourant à la méthode K-means, un algorithme de classification automatique basé sur la technique d'apprentissage non supervisé, inspiré de l'Intelligence Artificielle ;
- apprécier leur fiabilité qui est tributaire de la qualité des approches et outils utilisés pour produire ces résultats ;
- choisir le meilleur « classifieur » à recommander à cette IMF pour, d'une part, l'aider à ne consentir des prêts qu'à des candidats réellement crédibles, afin de minimiser le risque de contrepartie encouru, et d'autre part, déterminer les montants et délais de crédit à leur consentir par rapport à ceux sollicités.

Après analyse, il s'est avéré que le résultat de segmentation issu de la pratique en vigueur au sein de l'IMF ALPHA pour les nouveaux clients a été considéré comme sujet à caution et donc peu fiable. Il en est de même pour celui produit par le logiciel de scoring utilisé pour les anciens clients. Par contre celui produit par la méthode K-means comporte plusieurs atouts qui plaident en sa faveur et qui confirment son niveau élevé de fiabilité. C'est ainsi qu'on a recommandé à l'IMF ALPHA de privilégier cette dernière approche pour fiabiliser son processus d'octroi des crédits en affectant tout prétendant à un prêt, qu'il s'agisse d'un ancien ou d'un nouveau client, dans la classe d'appartenance correspondant à son degré réel de crédibilité. Une démarche logique a été proposée à l'IMF ALPHA au § 4.3. pour la guider dans la mise en œuvre de ce processus.

NOTES BIBLIOGRAPHIQUES

A. OUVRAGES ET ARTICLES DES REVUES

1. ALEXANDRE L., *La Guerre des intelligences. Intelligence artificielle versus intelligence humaine*, Paris, JC Lattès, 2017, 250 p. (ISBN 978-2-7096-6084-6, lire en ligne [archive])
2. ALLIOT J.-M. et SCHIEX T., *Intelligence artificielle et informatique théorique*, Toulouse, CEPADUES, 2002, 543 p. (ISBN 2-85428-578-6)
3. BARAK, N.M., *La Microfinance : Une solution de financement pour l'Afrique*, 2014
4. BENZECRI, J.P. et Coll., *Analyse des Données*, 2 vol., Dunod, 1973
5. BERSINI H., *De l'intelligence humaine à l'intelligence artificielle*, Paris, Ellipse, 2006, 192 p. (ISBN 2-7298-2813-3)

6. BOSS G., *Les machines à penser : L'homme et l'ordinateur*, Zurich, Éditions du Grand midi, 1987, 202 p. (ISBN 2-88093-105-3)
7. CAILLIEZ, F., ESCOUFIER, Y., *Analyse factorielle : un peu d'histoire et de géométrie*. Revue de Statistique Appliquée, Vol XXVII, n°1 pp. 5-28., 1979
8. CAZES, P., (1980). *L'analyse de certains tableaux rectangulaires décomposés en blocs : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. II Questionnaire : variantes de codages et nouveaux calculs de contributions*, Cahiers de l'Analyse des données, Vol 5 n°4 pp. 387-403.).
9. CHALLONER J. (trad. de l'anglais), *L'Intelligence artificielle : Un guide d'initiation au futur de l'informatique et de la robotique*, Paris, Pearson Education, 2003, 72 p. (ISBN 2-7440-1600-4)
10. CHAUDET H. et PELLEGRIN L., *Intelligence artificielle et psychologie cognitive*, Paris, Dunod, 1998, 179 p. (ISBN 2-10-002989-4)
11. CHAVENT M., GUINOT C., LECHEVALLIER Y., TENENHAUS M., *Méthodes divisives de classification et segmentation non supervisée : recherche d'une typologie de la peau humaine saine*, Revue de Statistiques Appliquées, XLVII (4), 87 – 99, 1999.
12. CHEVALIER F. et LE BELLAC J., *La classification*, Faculté des Sciences Economiques, Université de Rennes, 2013
13. COBAC, *Recueil des textes relatifs à l'exercice des activités de microfinance*. Secrétariat Général, Département microfinance, 2002
14. CREMILLEUX B., *Classification Interactive, Apprentissage par l'interaction*, Edition Europa, 1997.
15. CREVIER D. et BUCKCEK N. (trad. de l'anglais), *À la recherche de l'intelligence artificielle*, Paris, Flammarion, 1997, 438 p. (ISBN 2-08-081428-1), (traduction de (en) *The Tumultuous history of the search for artificial intelligence.*)
16. DE LAGARDE J., *Initiation à l'Analyse des Données*, 3^{ème} édition, Dunod, Paris, 1995, p. 28)
17. DELAHAYE J.-P., *Outils logiques pour l'intelligence artificielle*, Eyrolles, 1985 [détail des éditions] (ISBN 978-2212084122)
18. DEMSAR, J. & ZUPAN, BLAZ (2013). *Orange: Data mining fruitful and fun - A historical perspective*. Informatica (Slovenia). 37. 55-60.
19. DIDAY Edwin et al., *Éléments d'analyse de données*, 3^{ème} édition, Paris, Edition Bordas, 1982.
20. DREYFUS G., MARTINEZ J.M., SAMUELIDES M, GORDON M.B., BODRAN F., THIRIA S. : *Apprentissage statistique, Réseaux de Neurones ; Cartes Topologiques, Machines à vecteurs supports ;* Eyrolles, Paris, 2007
21. DUCIMETIERE, P., *Les méthodes de la classification numérique*, Revue de statistique appliquée, tome 18, n°4 (1970), p.5-25, (https://www.numdam.org/item?id=RSA_1970_8_4_29_0)
22. ERCOLE PALMERI (août 2020) , *Classification des algorithmes d'apprentissage automatique : régression linéaire, classification et clustering*,
23. ESCOFIER, B., PAGES, J., *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*, Dunod, 1988.
24. FOUCART Thierry, *L'analyse des données. Mode d'emploi*, Rennes, Presses Universitaires de Rennes, 1997.
25. GANASCIA J.-G., *L'Intelligence artificielle*, Paris, Éditions du Cavalier bleu, coll. « Idées recues », 2007, 127 p. (ISBN 978-2-84670-165-5)
26. GHEWY Pierre, *Guide pratique de l'analyse des données*, Bruxelles, De Boeck, 2010.
27. HAND D., MANILLA H., SMYTH P., *Principles of data mining*, Bardford Books, 2001.
28. HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., *Les Eléments de l'Apprentissage Statistique. Exploration des Données, Inférence et Prédiction* (deuxième édition), Springer, 2001
29. HATON J.-P. , HATON M.-C., *L'Intelligence Artificielle*, Paris, Que sais-je ?, 1990, 127 p. (ISBN 2-13-043164-X)
30. HUNT E.B., *Concept Learning: An Information Processing Problem*, Wiley, 1962.
31. KAFUNDA KATALAY, P., *Contribution à la Segmentation dans un Environnement Incertain Basée sur l'Apprentissage non Déterministe et sur le Système Neuro-Flou Hybride : Application au syndrome Métabolique*, Thèse de Doctorat, Faculté des Sciences, Département de Mathématique et Informatique, Unikin, 2013
32. KASORO MULENDA Nathanaël, *Data Mining et Analyse des Données : Pour une Nouvelle Méthode de Classification Automatique des Données Symboliques Intervalles, Basée sur les Processus de Poisson*

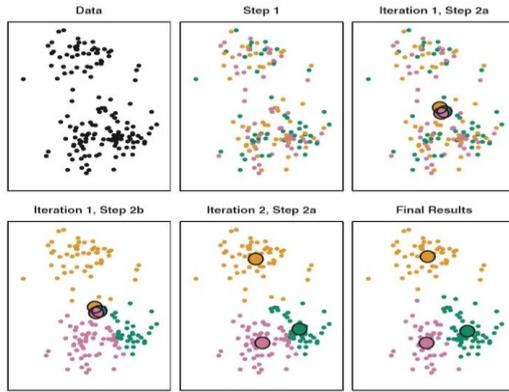
- Homogènes*, Thèse de doctorat en Sciences Informatiques, Université de Kinshasa, 2009.
33. KASS G., (1980), « *An exploratory Technique for Investigating Large Quantities of Categorical Data* », *Applied Statistics*, 29(2), 119-127.
 34. KOHAVI R., SOMMERFIED D., *Handbook of Data Mining and Knowledge Discovery*, In Will Klossgen and Jan M. Zytkow, editors, , chapter 24.1.2, pages 548-553. Oxford University Press, 2002
 35. LEBART L., MORINEAU A., PIRON M., *Statistique Exploratoire Multidimensionnelle*, 3^{ème} édition, Dunod, 2000
 36. LEFEVRE Jacques, *Introduction aux analyses statistiques multidimensionnelles*, Masson, Paris, 1976.
 37. MANDIANGU N. : « *Contribution à la définition d'un cadre de norme juridique spécifique pour la néo-microfinance aux Comores*. Thèse de doctorat en Droit d'économie et Finance. Université Internationale Bircham, 2012.
 38. MAPHANA ma NGUMA S., *Méthodes Quantitatives de Gestion en Marketing : Soubassement Théorique*, Séminaire de DEA 2, Faculté des Sciences Economiques et de Gestion, Unikin, 2019.
 39. MAYOUKOU (2015), « *Les banques multinationales et l'offre de services en microfinance : peut-on parler d'éthique microfinancière ? Une analyse par le paradigme « OLI » revisité* », in *Ethique, entrepreneuriat et développement*, Les cahiers de l'Association Tiers-Monde, n° 30, 2015
 40. MOUALEK Y., *Deep Learning pour la Classification des Images*, Mémoire de Master en Informatique, Université Abou Beckr Belkaid, Alger 2017
 41. NTAMBWE KABAMBA. C., *Etude de la Reconnaissance des formes par l'Approche Deep Learning*, mémoire de DEA en Sciences Informatiques, Faculté des Sciences, Département des Mathématiques et Informatique, Université de Kinshasa, 2021.
 42. PITRAT J. , *Artificial Beings, the conscience of a conscious machine*, ISTE-Wiley, 2019 (ISBN 9781848211018).
 43. QUINLAN R., *C4.5: Programs for Machine Learning*, Morgan Kaufman, 1993.
 44. RAKOTOMALALA R., *Méthodes des centres mobiles. Classification par partition. Les méthodes de réallocation*.
<http://tutoriels-data-mining.blogspot.fr/>
 45. RAKOTOMALALA R., TANAGRA : *Une Plate-Forme d'Expérimentation pour la Fouille de Données*", *Revue MODULAD*, 32, 70-85, 2005.
 46. RAO, R.,C., (1948) : *L'utilisation des Mesures Multiples dans les Problèmes de Classification Biologique*, *Journal de la Royal Statistic Society*, série B. 10(2), 159-203. JSTOR 2983775
 47. RUSSEL S. J. et NORVIG P. , *Intelligence artificielle (3^e éd.)* [« *Artificial Intelligence: A Modern Approach* (en) »], Pearson Education France, 2010, 1198 p. (ISBN 978-2-7440-7455-4, présentation en ligne [archive]).
 48. SABAH G. , *L'Intelligence artificielle et le langage, Représentations des connaissances, Processus de compréhension*, vol. 1, Hermès, 1989 (ISBN 2-86601-134-1)
 49. SABAH G., *L'Intelligence artificielle et le langage, Représentations des connaissances, Processus de compréhension*, vol. 2, Paris, Hermès, 1990, 768 p. (ISBN 2-86601-187-2)
 50. SABAH G., *Compréhension des langues et interaction (Traité IC2, Série Cognition et Traitement de l'Information)*, Paris, Hermès science: Lavoisier, 2006, 400 p. (ISBN 2-7462-1256-0)
 51. SAPORTA G., « *Probabilités, Analyse des Données et Statistique*, 2^{ème} édition, Technip, 2006
 52. SERVET, J.M.. *La vraie révolution du microcrédit*. Edition Odile Jacob économie, Paris, 2015
 53. STAFFORD Jean et BODSON Paul, *L'analyse multi-variée avec SPSS*, Presses de l'Université du Québec, Québec, 2007.
 54. STARR, M.K. et MILLER D.M., *La gestion des stocks : théorie et pratique*, Dunod, Paris, 1966.
 55. TISSEAU G. et PITRAT J., *Intelligence artificielle : problèmes et méthodes*, Paris, Presses universitaires de France, 1996, 255 p. (ISBN 2-13-047429-2)
 56. TOMASONE, R., *Analyse multidimensionnelle et classification*, *Revue de statistique appliquée*, tome 18, n°4 (1970), p.29-34,
https://www.numdam.org/article/RSA_1970_8_4_29_0
 57. TOMASONE R., DANZART M., DAUDIN J. J., MASSON J.P., *Discrimination et classement*, Masson, 1988
 58. TUFFERY S., *Etude de cas en statistique décisionnelle*, Technip, 2009
 59. TUFFERY S., *Data mining et statistique décisionnelle*, Technip, 2010
 60. TURING Alan, GIRARD Jean-Yves, *La Machine de Turing*, Éditions du Seuil, 1995 [détail de

l'édition], *Les Ordinateurs et l'Intelligence*, p. 133-174

61. VENTRE D., *Intelligence artificielle, cybersécurité et cyberdéfense*, ISTE, Londres, 246 pages, juin 2020 (ISBN papier : 9781784056797), (ISBN ebook : 9781784066796)
 62. VOLLE M., *Analyse des données*, 3^{ème} édition, Economica, 1993.
 63. WITTEN I., FRANCK E., *Data Mining: Practical machine learning tools with Java implementations*, Morgan Kaufmann, San Francisco, 2000.
 64. YERE, I. (2017). *Pérennité des Institutions de Microfinance en République Démocratique du Congo : Une application de l'approche analytique du trinôme garant de la performance*. Thèse de doctorat, 419, 328-329
- B. AUTRES DOCUMENTS**
65. *Rapport annuel*, Banque Centrale du Congo (BCC), 2019
 66. *Loi fixant les règles relatives à l'activité de microfinance en République*, Journal officiel de la RDC, 2011
 67. *Rapport sur le développement dans le monde*, Banque Mondiale, 1998.
 68. *Stratégie de la Banque Mondiale pour le développement de la microfinance et des services financiers aux PME et en milieu rural en Afrique subsaharienne*, Banque Mondiale, Rapport, n° 21202, Washington DC, août 1998
- C. WEBOGRAPHIE**
69. <https://www.talend.com/fr/ressources/methode-modele-data-mining>
 70. LIAUDET B., Cours de Data Mining 1 : Introduction, 2008, <https://studylibfr.com/doc/3085558/cours-de-data-mining-1---introduction>
 71. LIAUDET B., Cours de Data Mining 3 : Modélisation, Présentation générale, 2008 <https://studylibfr.com/doc/2264223/cours-de-data-mining-3---site-de-bertrand-liaudet>
 72. <https://www.futura-sciences.com/tech/actualites/informatique-interview-nee-intelligence-artificielle-90946>
 73. <https://experience.microsoft.fr/articles/intelligence-artificielle/apprentissage-supervise-et-non-supervise-quelles-differences/>
 74. <https://analyticsinsights.io/quest-ce-que-machine-learning-importnat/>
 75. <https://techno-science.net/glossaire-definition/apprentissage-automatique.html>
 76. <https://datascience.eu.fr/apprentissage-automatique/quest-ce-que-l-apprentissage-automatique-une-definition>
 77. www.psychomedia.qc.ca/lexique/definition/apprentissage_automatique
 78. <https://www.ibm.com/fr-fr/cloud/learn/machine-learning?msclkid=97d9c03fab8d11eca8ebf8ebbc70ea0d>
 79. <https://datascientest.com/transfer-learning>
 80. <https://www.ibm.com/fr-fr/cloud/analytics/machine-learning?msclkid=3afdca22ab8811ecaa9b535>
 81. <https://machinelearnig.com/apprentissage-supervise-4-etapes/>
 82. <https://cedric.cnam.fr/vertigo/ml/CoursClassificationAutomatic.html>
 83. <http://iml.univ-mrs.fr>
 84. <https://medium.com/tell-ia/a-regression-logistique-expliquee-a-ma-grand-mere>
 85. www.microfinance.sn/la-microfinance/
 86. <https://www.ladissertation.com/Sciences-Economiques-et-Sociales...>
 87. <https://www.cscience.ca/2021/9/analyse-ia-forte-versus-ia-faible-pourquoi-ce-rapport-de-force/>

ANNEXES

ANNEXE 1 : Illustration des diagrammes de visualisation des résultats fournis par K-Means



Ces graphiques sont tirées de l'ouvrage 'An introduction to statistical learning, de James, G., Witten, D., Hastie, T., & Tibshirani, (2013)'

Fig. 1. Exemple illustrant le processus de classification des individus selon l'algorithme K-means [82]

Dans cet exemple, la solution obtenue optimale est atteinte après 5 itérations.

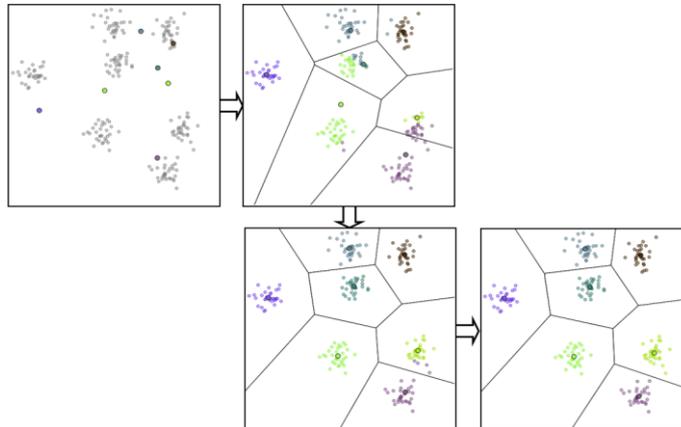


Fig. 2. Illustration des étapes successives dans l'application de K-means (résultat final : 7 classes) [82]

Dans cet exemple, la solution obtenue optimale est atteinte après 2 itérations seulement.

ANNEXE 2 : Données caractérisant les 100 candidats emprunteurs retenus dans l'échantillon

Clients	(R ₁)	(R ₃)	(R ₅)	(R ₆)	(R ₇)	(R ₉)	(R ₁₀)	(R ₁₁)	Montant (\$)		Délai (mois)		Acompte mens.		Class. ALPHA
									Sollicité	Accordé	Sollicité	Accordé	Sollicité	Accordé	
C001	0,25	0,38	0,71	2,89	1,00	0,14	1,03	TH	60 000	60 000	18	18	3 333	3 333	C3
C002	0,25	0,25	0,60	4,72	0,99	0,33	0,26	TH	50 000	50 000	20	20	2 500	2 500	C3
C003	0,02	0,39	2,23	1,68	4,52	0,25	0,13	PH	50 000	30 000	24	15	2 083	2 000	C1
C004	0,17	0,23	0,33	11,40	1,11	0,25	0,75	PH	30 000	19 000	24	12	1 250	1 583	C1
C005	0,06	0,34	1,21	3,55	1,43	0,25	0,50	PH	25 000	20 000	20	18	1 250	1 111	C1
C006	0,41	0,41	0,41	3,67	1,03	0,45	0,20	HO	15 000	13 000	12	12	1 250	1 083	C2
C007	0,45	0,20	0,32	6,18	1,04	0,07	2,10	HO	15 000	12 000	12	12	1 250	1 000	C2
C008	0,30	0,26	0,39	3,77	0,23	0,36	0,45	PH	12 000	9 000	12	10	1 000	900	C1
C009	0,10	0,38	0,35	7,40	1,45	0,19	1,12	PH	15 000	8 000	15	10	1 000	800	C1
C010	0,17	0,33	1,92	1,36	0,50	0,25	0,60	PH	10 000	7 000	12	10	833	700	C1
C011	0,36	0,18	0,44	3,72	0,35	0,22	1,30	HO	10 000	7 000	12	12	833	583	C2
C012	0,31	0,33	0,28	7,95	0,18	0,36	0,28	PH	30 000	7 000	24	10	1 250	700	C1
C013	0,27	0,35	0,40	7,34	0,60	0,28	0,53	TH	5 000	5 000	12	12	416	416	C3
C014	0,01	0,30	0,40	4,66	0,27	0,28	0,40	HO	5 000	4 500	8	8	625	562	C2
C015	0,48	0,39	1,87	2,96	0,70	0,14	1,87	HO	15 000	4 000	12	12	1 250	333	C2
C016	0,34	0,17	0,41	4,06	0,47	0,29	0,86	PH	5 000	4 000	12	10	416	400	C1

C017	0,30	0,18	0,40	4,74	0,59	0,32	0,75	HO	4 000	3 800	10	10	400	380	C2
C018	0,27	0,23	0,43	6,72	0,50	0,43	0,52	HO	5 000	3 000	12	12	416	250	C2
C019	0,40	0,23	0,69	2,16	0,41	0,19	1,97	PH	2 500	2 000	10	8	250	250	C1
C020	0,26	0,30	0,44	6,63	0,67	0,30	0,77	HO	4 500	2 150	12	12	375	179	C2
C021	0,37	0,33	0,64	3,34	1,86	0,18	3,23	HO	5 000	2 000	8	8	625	250	C2
C022	0,40	0,31	0,65	2,79	0,40	0,19	1,33	PH	5 000	3 000	10	8	500	375	C1
C023	0,21	0,07	0,37	7,80	0,53	0,21	1,67	HO	2 500	2 000	12	12	208	166	C2
C024	0,47	0,21	0,68	1,93	0,37	0,29	1,33	PH	3 000	2 000	12	8	250	250	C1
C025	0,21	0,20	0,62	3,43	0,60	0,25	12,99	HO	2 000	2 000	8	6	250	333	C3
C026	0,39	0,28	0,41	4,21	0,48	0,21	1,28	PH	2 500	1 950	10	8	250	240	C1
C027	0,18	0,26	0,25	8,39	0,53	0,38	0,50	HO	1 800	1 800	12	8	150	225	C3
C028	0,24	0,37	0,46	5,93	0,11	0,11	3,10	HO	5 000	1 500	12	8	416	187	C1
C029	0,30	0,23	0,67	3,80	0,68	0,20	2,43	TH	1 500	1 500	10	10	150	150	C3
C030	0,35	0,33	0,40	1,34	0,61	0,22	1,30	PH	2 500	1 500	10	8	250	187	C1
C031	0,31	0,27	0,40	5,38	0,97	0,20	1,26	TH	1 500	1 500	8	8	187	187	C3
C032	0,19	0,26	0,46	4,95	1,45	0,24	1,72	PH	3 000	1 500	10	8	300	187	C1
C033	0,22	0,23	0,74	2,89	1,25	0,21	2,42	HO	1 500	1 500	8	7	187	214	C3
C034	0,26	0,33	0,45	4,73	1,00	0,23	1,15	HO	2 000	1 500	8	8	250	187	C2
C035	0,18	0,28	0,33	6,48	1,07	0,21	1,18	HO	1 500	1 300	8	8	187	162	C2
C036	0,23	0,33	0,42	5,06	1,08	0,29	1,01	HO	1 500	1 250	8	8	187	156	C2
C037	0,21	0,36	0,39	5,36	0,70	0,23	1,39	HO	2 500	1 000	8	8	312	125	C2
C038	0,22	0,10	0,48	5,65	1,07	0,23	1,30	HO	1 000	1 000	12	10	83	100	C3
C039	0,21	0,27	0,49	1,69	1,14	0,23	1,22	TH	1 000	1 000	10	10	100	100	C3
C040	0,20	0,21	0,50	4,61	0,54	0,19	2,02	TH	1 000	1 000	8	8	125	125	C3
C041	0,27	0,36	0,81	3,75	1,10	0,15	1,22	HO	80 000	80 000	17	16	4 705	5 000	C3
C042	0,23	0,22	0,70	3,60	1,15	0,34	0,29	HO	70 000	70 000	20	18	3 500	3 888	C3
C043	0,40	0,40	4,60	1,70	4,81	0,44	0,15	PH	70 000	50 000	24	20	2 916	2 500	C1
C044	0,20	0,41	0,43	12,60	2,21	0,37	0,85	PH	40 000	28 000	24	20	1 666	1 400	C1
C045	0,08	0,40	4,24	3,22	1,75	0,48	1,30	PH	30 000	20 000	20	18	1 500	1 111	C1
C046	0,43	0,44	0,51	4,23	1,57	0,47	3,18	TH	25 000	25 000	12	12	2 083	2 083	C3
C047	0,47	0,21	0,41	6,20	1,23	0,08	3,20	HO	30 000	20 000	12	12	2 500	1 666	C2
C048	0,17	0,33	33,85	1,74	0,11	0,07	2,15	PH	15 000	8 000	12	10	1 250	800	C1
C049	0,11	0,39	0,46	8,40	1,53	0,23	1,18	PH	30 000	27 000	15	14	2 000	1 928	C1
C050	0,18	0,34	36,60	1,27	1,47	0,02	1,28	PH	25 000	19 000	12	10	2 083	1 900	C1

Clients	(R ₁)	(R ₃)	(R ₅)	(R ₆)	(R ₇)	(R ₉)	(R ₁₀)	(R ₁₁)	Montant (\$)		Délai (mois)		Acompte mens.		Class. ALPHA
									Sollicité	Accordé	Sollicité	Accordé	Sollicité	Accordé	
C051	0,26	0,32	0,47	6,65	0,69	0,31	0,81	PH	45 000	24 000	12	12	3 750	2 000	C2
C052	0,90	0,31	0,53	4,36	0,27	0,22	0,74	TH	10 000	10 000	12	12	833	833	C3
C053	0,27	0,24	0,43	6,81	0,51	0,43	0,64	PH	55 000	50 000	24	22	2 291	2 272	C1
C054	0,48	0,38	1,07	2,98	0,79	0,46	0,78	PH	40 000	32 000	24	15	1 666	2 133	C1
C055	0,34	0,18	0,50	4,76	0,87	0,29	0,83	PH	50 000	23 000	20	15	2 500	1 533	C1
C056	0,40	0,26	0,67	3,41	0,41	0,18	1,64	TH	25 000	25 000	10	10	2 500	2 500	C3
C057	0,50	0,18	0,54	4,71	0,58	0,35	1,78	HO	40 000	25 000	15	15	2 666	1 666	C2
C058	0,60	0,19	0,64	5,11	0,63	0,34	1,27	HO	60 000	55 000	24	24	2 500	2 291	C2
C059	0,27	0,36	0,51	4,20	0,60	0,88	0,51	HO	10 000	9 000	12	12	833	750	C2
C060	0,34	0,19	3,75	8,40	0,18	0,37	0,27	PH	30 000	12 000	24	11	1 250	1 090	C1

C061	0,00	0,38	0,68	1,38	0,66	0,19	2,75	HO	25 000	20 000	16	16	1 562	1 250	C2
C062	0,30	0,23	0,70	3,90	0,71	0,19	2,93	PH	35 000	30 000	18	16	1 944	1 875	C1
C063	0,25	0,38	0,48	5,72	0,67	0,13	4,30	HO	50 000	40 000	15	15	3 333	2 666	C2
C064	0,37	0,03	0,65	3,47	1,87	0,18	3,24	HO	20 000	20 000	16	14	1 250	1 428	C3
C065	0,01	0,31	0,67	2,81	0,40	0,19	1,33	HO	7 000	5 000	12	12	583	416	C2
C066	0,23	0,07	0,35	7,90	0,53	0,21	1,33	HO	7 500	6 500	12	12	625	541	C2
C067	0,48	0,22	0,67	1,93	0,37	0,29	1,34	PH	3 500	1 500	12	8	291	187	C1
C068	0,21	0,20	0,64	3,22	0,61	0,26	12,29	HO	3 000	3 000	8	6	375	500	C3
C069	0,41	0,27	0,42	4,15	0,52	0,21	1,27	PH	6 000	4 000	10	8	600	500	C1
C070	0,23	0,87	0,25	9,34	0,58	0,37	0,53	PH	3 000	2 500	12	8	250	312	C1
C071	0,34	0,29	0,45	6,03	0,98	0,21	1,31	TH	15 000	15 000	20	20	750	750	C3
C072	0,95	0,27	0,47	4,17	1,45	0,23	1,74	HO	65 000	35 000	20	20	3 250	1 750	C2
C073	0,29	0,24	0,74	2,87	1,27	0,21	2,50	HO	15 000	15 000	15	14	1 000	1 071	C3
C074	0,25	0,34	0,45	4,63	1,11	0,23	1,15	TH	20 000	20 000	15	15	1 333	1 333	C3
C075	0,18	0,24	0,35	6,49	1,08	0,21	1,71	TH	15 000	15 000	15	15	1 000	1 000	C3
C076	0,24	0,33	0,45	7,07	1,10	0,28	1,01	HO	16 000	16 000	18	16	888	1 000	C3
C077	0,24	0,37	0,44	5,45	0,76	0,24	1,49	HO	25 000	25 000	16	15	1 562	1 666	C3
C078	0,23	0,10	0,37	5,65	1,07	0,23	2,64	HO	10 000	10 000	12	11	833	909	C3
C079	0,25	0,10	0,48	5,04	1,16	0,22	1,71	PH	40 000	35 000	18	15	2 222	2 333	C1
C080	0,23	0,21	0,74	4,65	0,54	0,19	2,04	PH	11 000	10 000	14	12	785	833	C1
C081	0,26	0,48	0,81	2,79	1,12	0,16	1,07	HO	70 000	67 000	20	20	3 500	3 350	C2
C082	0,23	0,25	0,80	4,73	1,06	0,34	0,25	HO	80 000	75 000	24	24	3 333	3 125	C2
C083	0,08	0,39	0,94	2,25	1,53	0,45	0,15	PH	75 000	68 000	22	20	3 409	3 400	C1
C084	0,18	0,22	0,44	11,80	1,16	0,27	0,78	PH	35 000	20 000	24	12	1 458	1 666	C1
C085	0,07	0,34	1,24	3,74	1,45	0,35	1,04	PH	30 000	25 000	15	12	2 000	2 083	C1
C086	0,43	0,42	0,43	3,77	1,09	0,55	0,30	HO	16 000	15 000	12	12	1 333	1 250	C2
C087	0,44	0,21	0,34	7,30	1,08	0,08	3,01	HO	15 000	12 000	12	12	1 250	1 000	C2
C088	0,71	0,26	0,37	3,75	0,25	0,36	0,74	PH	14 000	8 000	12	10	1 166	800	C1
C089	0,11	0,36	0,37	7,80	1,45	0,20	1,14	PH	17 000	12 000	12	10	1 416	1 200	C1
C090	0,18	0,34	1,25	1,37	0,01	0,28	0,17	PH	10 000	4 000	12	10	833	400	C1
C091	0,38	0,01	0,45	3,74	0,40	0,24	1,80	PH	15 000	10 000	15	12	1 000	833	C1
C092	0,32	0,25	0,30	7,95	0,18	0,36	0,33	PH	35 000	20 000	16	14	2 187	1 428	C1
C093	0,28	0,54	0,45	7,38	0,60	0,30	0,53	PH	10 000	9 000	12	10	833	900	C1
C094	0,08	0,63	0,46	4,67	0,28	0,29	0,75	PH	7 000	5 000	12	10	583	500	C1
C095	0,47	0,01	1,78	1,69	0,63	0,44	1,85	PH	38 000	20 000	20	15	1 900	1 333	C1
C096	0,36	0,87	0,45	4,07	0,47	0,33	0,87	HO	10 000	8 000	12	12	833	666	C2
C097	0,40	0,76	0,50	5,64	0,50	0,36	0,85	HO	7 000	6 000	12	12	583	500	C2
C098	0,28	0,58	0,43	6,72	0,59	0,43	0,64	PH	16 000	9 000	15	14	1 066	642	C1
C099	0,45	0,02	0,87	3,14	0,46	0,19	1,87	PH	25 000	15 000	20	15	1 250	1 000	C1
C100	0,27	0,97	0,71	0,74	0,45	0,29	0,75	HO	4 500	4 000	12	12	375	333	C2
Norme	<0,50	<0,70	≤1,00	≥1,00	≥0,80	≥0,20	>1,00	≥PH							

Source : Résultats d'enquête auprès de l'IMF ALPHA

Annexe 3. : Evaluations discrétisées des demandeurs des crédits

Clients	(R ₁)	(R ₂)	(R ₃)	(R ₄)	(R ₅)	(R ₆)	(R ₇)	(R ₈)	(R ₉)	Class.ALPHA	Class.K-means	(*)
C001	3	2	2	1	2	1	2	4		C3	C2	
C002	3	3	3	2	2	2	1	4		C3	C3	
C003	4	2	1	1	4	2	1	1		C1	C1	
C004	3	3	4	4	2	2	1	1		C1	C3	(*)
C005	4	3	1	1	4	2	1	1		C1	C1	

C006	1	2	4	1	2	3	1	2	C2	C3	
C007	1	3	4	2	2	1	3	2	C2	C3	
C008	2	3	4	1	1	2	1	1	C1	C3	(*)
C009	4	2	4	3	4	1	2	1	C1	C3	(*)
C010	3	3	1	1	1	2	1	1	C1	C1	
C011	2	4	4	1	1	1	2	2	C2	C2	
C012	2	3	4	3	1	2	1	1	C1	C3	(*)
C013	2	3	4	3	1	2	1	4	C3	C3	
C014	4	3	4	2	1	2	1	2	C2	C3	
C015	1	2	1	1	1	1	3	2	C2	C2	
C016	2	4	4	2	1	2	1	1	C1	C3	(*)
C017	2	4	4	2	1	2	1	2	C2	C3	
C018	2	3	4	2	1	2	1	2	C2	C3	
C019	1	3	2	1	1	1	3	1	C1	C2	
C020	2	3	4	2	1	2	1	2	C2	C3	
C021	2	3	2	1	4	1	4	2	C2	C2	
C022	1	3	2	1	1	1	2	1	C1	C2	
C023	3	4	4	3	1	1	2	2	C2	C3	
C024	1	3	2	1	1	2	2	1	C1	C2	
C025	3	3	2	1	1	2	4	2	C3	C2	
C026	1	3	4	2	1	1	2	1	C1	C3	(*)
C027	3	3	4	3	1	2	1	2	C3	C3	
C028	3	2	3	2	1	1	4	2	C1	C2	
C029	2	3	2	1	1	1	3	4	C3	C2	
C030	2	3	4	1	1	1	2	1	C1	C2	
C031	2	3	4	2	1	1	2	4	C3	C2	
C032	3	3	3	2	4	2	2	1	C1	C3	(*)
C033	3	3	2	1	3	1	3	2	C3	C2	
C034	2	3	3	2	2	1	2	2	C2	C2	
C035	3	3	4	2	2	1	2	2	C2	C3	
C036	2	3	4	2	2	2	2	2	C2	C3	
C037	3	3	4	2	1	1	2	2	C2	C3	
C038	3	4	3	2	2	1	2	2	C3	C2	
C039	3	3	3	1	2	1	2	4	C3	C2	
C040	3	3	3	2	1	1	3	4	C3	C2	
C041	2	3	2	1	2	1	2	2	C3	C2	
C042	3	3	2	1	3	2	1	2	C3	C1	
C043	1	2	1	1	4	2	1	1	C1	C1	
C044	3	2	4	4	4	2	1	1	C1	C3	(*)
C045	4	2	1	1	4	3	2	1	C1	C1	
C046	1	2	3	2	4	3	4	4	C3	C2	
C047	1	3	4	2	4	1	4	2	C2	C2	
C048	3	3	1	1	1	1	3	1	C1	C2	
C049	4	2	3	3	4	1	2	1	C1	C3	(*)
C050	3	3	1	1	4	1	2	1	C1	C1	

Clients	(R ₁)	(R ₂)	(R ₃)	(R ₄)	(R ₅)	(R ₆)	(R ₇)	(R ₈)	(R ₉)	Class.ALPHA	Class.K-means	(*)
C051	2	3	3	2	1	2	1	1		C2	C3	
C052	1	3	3	2	1	1	1	4		C3	C2	
C053	2	3	4	3	1	2	1	1		C1	C3	(*)
C054	1	2	1	1	1	3	1	1		C1	C1	
C055	2	4	3	2	1	2	1	1		C1	C3	(*)

C056	1	3	2	1	1	1	2	4	C3	C2	
C057	1	4	2	2	1	2	3	2	C2	C2	
C058	1	4	2	2	1	2	2	2	C2	C2	
C059	2	3	4	2	1	4	1	2	C2	C3	
C060	2	4	1	3	1	2	1	1	C1	C3	(*)
C061	4	2	2	1	1	1	4	2	C2	C2	
C062	2	3	2	2	1	1	4	1	C1	C2	
C063	3	2	3	2	1	1	4	2	C2	C2	
C064	2	4	2	1	4	1	4	2	C3	C2	
C065	4	3	2	1	1	1	2	2	C2	C2	
C066	3	4	4	3	1	1	2	2	C2	C3	
C067	1	3	4	1	1	2	2	1	C1	C3	(*)
C068	3	3	2	1	1	2	4	2	C3	C2	
C069	1	3	4	2	1	1	2	1	C1	C3	(*)
C070	3	1	4	3	1	2	1	1	C1	C3	(*)
C071	2	3	3	2	2	1	2	4	C3	C2	
C072	1	3	3	2	4	1	2	2	C2	C2	
C073	2	3	2	1	3	1	4	2	C3	C2	
C074	3	3	3	2	2	1	2	4	C3	C2	
C075	3	3	4	2	2	1	2	4	C3	C2	
C076	3	3	3	3	2	2	2	2	C3	C3	
C077	3	2	4	2	1	2	2	2	C3	C3	
C078	3	4	4	2	2	1	4	2	C3	C2	
C079	3	4	3	2	3	1	2	1	C1	C2	
C080	3	3	2	2	1	1	3	1	C1	C2	
C081	2	2	2	1	2	1	2	2	C2	C2	
C082	3	3	2	2	2	2	1	2	C2	C3	
C083	4	2	1	1	4	3	1	1	C1	C1	
C084	3	3	4	4	3	2	1	1	C1	C3	(*)
C085	4	3	1	1	4	2	2	1	C1	C1	
C086	1	2	4	1	2	3	1	2	C2	C3	(*)
C087	1	3	4	3	2	1	4	2	C2	C2	
C088	1	3	4	1	1	2	1	1	C1	C3	(*)
C089	4	3	4	3	4	1	2	1	C1	C3	(*)
C090	3	3	1	1	1	2	1	1	C1	C1	
C091	2	4	3	1	1	2	3	1	C1	C2	
C092	2	3	4	3	1	2	1	1	C1	C3	(*)
C093	2	1	3	3	1	2	1	1	C1	C3	(*)
C094	4	1	3	2	1	2	1	1	C1	C3	(*)
C095	1	4	1	1	1	3	3	1	C1	C1	
C096	2	1	3	2	1	2	1	2	C2	C3	
C097	1	1	3	2	1	2	1	2	C2	C3	
C098	2	1	4	2	1	2	1	1	C1	C3	(*)
C099	1	4	1	1	1	1	3	1	C1	C2	
C100	2	1	2	1	1	2	1	2	C2	C1	

Source : élaboré sur base des données du Tableau 2 (pour les évaluations discrétisées), de l'Annexe 2 (pour la classification selon l'IMF ALPHA) ainsi que du logiciel ORANGE DATA (pour la classification K-means.