



المندوبية السامية للتخطيط
HAUT-COMMISSARIAT AU PLAN

ROYAUME DU MAROC

*_*_*_*_*

HAUT COMMISSARIAT AU PLAN

*_*_*_*_*_*_*_*

INSTITUT NATIONAL

DE STATISTIQUE ET D'ECONOMIE APPLIQUEE

INSEA



Projet de Fin d'Etudes

Elaboration d'une grille de score pour les TPE

-Crédit Automobile-

Préparé par : *M.ATMANI Youssef*
Mlle. LOUKILI Dounia

Sous la direction de : *M. MARRI Fouad (INSEA)*

M.BELLOUELJA Abdellatif (WAFASALAF)

Soutenu publiquement comme exigence partielle en vue de l'obtention du

Diplôme d'Ingénieur d'Etat

Option : Actuariat- Finance

Devant le jury composé de :

- *M.MARRI Fouad (INSEA)*
- *M. DOGHMI Ahmed (INSEA)*
- *M.BELLOUELJA Abdellatif (WAFSALAF)*

Résumé

Face à la conjoncture actuelle et aux faibles rendements engendrés par une crise qui a impacté tous les domaines, au Maroc notamment. Les institutions financières se mobilisent pour une meilleure gestion des risques. C'est dans cette perspective que l'établissement d'un système de notation interne, conformément à Bale II, relève des priorités au sein de chaque organisme.

Dans le présent rapport, nous avons étudié la faisabilité du projet pour ainsi arriver à établir une grille de notation adaptée aux très petites entreprises (TPE) pour les crédits automobile.

Pour ce faire, il a été nécessaire d'établir une méthodologie à travers laquelle on a pu achever notre travail. Ainsi, chaque chapitre du rapport comporte deux grands titres, le premier retrace le cadre théorique de chaque partie et le second concerne l'application de ce cadre théorique en pratique.

Mots clés :

Bale II,

Réglementation,

Risque crédit

, notation interne,

Credit scoring,

Défaut,

Régression logistique,

Classes de risque.

Dédicace

A celui qui m'a indiqué la bonne voie en me rappelant toujours que la volonté fait les grands hommes... A mon cher père ;

A celle qui a attendu avec patience le fruit de sa bonne éducation... A ma chère mère ;

A mon cher frère Mouncef et sa femme Hela ;

A ma chère Ghita ;

A mes chers amis : Seifeddine, Mehdi Akechtoul, Oussama, Walid, Reda, Naoufel, Mehdi Taib, Anas, Mansour, Alaa ;

A toute ma famille et à tous qui me sont chers ;

A mon binôme Dounia, sans qui ce travail n'aurait pas vu le jour.

Youssef Atmani

Dédicace

À mon très cher papa, qui a toujours su être présent pour moi, m'éclairer par ses conseils et me rappeler que la patience et la volonté sont les clés de la réussite.

À ma jolie maman, source de joie, de réconfort et de bonne humeur, merci d'être toujours là pour moi et m'entourer d'affection et d'amour.

À ma sœur Salma

À mes frères Moncef et Ziko

Je vous aime, Que dieu vous garde...

À mes grand-mères, mes tantes, mes oncles, mes cousins et cousines.

À la mémoire de ceux qui nous ont quittés...

À mes amis : Hajar, Mounia, Chaimae, Asmae, Seifeddine, Anas, Thyifa, Bouib, Amine et Hamza.

Vous occupez et vous occuperez toujours une place dans mon cœur.

À mon binôme Youssef, sans qui ce travail n'aurait pas vu le jour

Je vous dédie ce travail ...

Dounia Loukili

Remerciements

Tout d'abord nous tenons à exprimer nos remerciements à notre encadrant interne M. MARRI Fouad, pour nous avoir guidé par ses conseils et ses consignes et pour avoir été présent durant cette période de stage.

Nous tenons à remercier vivement notre encadrant externe M. BELLOUELJA Abdellatif pour nous avoir épaulé et guidé durant toute la période passée à Wafasalaf.

Nos remerciements s'adressent aussi à M. FEDDOUL Driss, pour nous avoir garanti toutes les conditions nécessaires au bon déroulement de ce PFE.

Nos sentiments de gratitude vont aussi à MM. OUAKIB Adil et. BOUYA Adil pour leur soutien et leur disponibilité.

Qu'il soit permis aussi de remercier l'ensemble du personnel de WAFASALAF pour son accueil et son professionnalisme notamment l'ensemble du service prévention risque et le service des engagements.

Nous remercions également le corps professoral de l'INSEA, pour leurs efforts et leur contribution dans l'épanouissement de l'institut.

Enfin, nous tenons à remercier toute personne ayant contribué de près ou de loin au bon déroulement de notre Projet de fin d'étude.

Table des matières

<i>Résumé et mots clés</i>	- 3 -
<i>Dédicace</i>	- 4 -
<i>Remerciements</i>	- 6 -
<i>Table des matières</i>	- 7 -
<i>Liste des abréviations</i>	10
<i>Liste des tableaux</i>	11
<i>Liste des figures</i>	13
<i>Chapitre 0 : Généralités et présentation du cadre de travail</i>	15
<i>I. Définitions et Généralités</i>	16
<i>I.1. Introduction</i>	16
<i>I.2. Définition risque de crédit</i>	16
<i>I.3. Ratio standards</i> :	17
<i>I.4. Définition du crédit Scoring</i> :	18
<i>I.5. Avantages et inconvénients du « Credit Scoring »</i> :	18
<i>I.6. Cas de Wafasalaf</i> :	19
<i>I.7. Secteur TPE au Maroc</i> :.....	19
<i>II. Présentation du cadre du travail</i>	20
<i>II.1. Préparation des données</i> :	20
<i>II.1.1. Critère et variable de défaut</i> :	20
<i>II.1.2. Fenêtre d'observation</i> :	20
<i>Chapitre 1 : Données et méthodologie de l'étude</i>	22
<i>I. Cadre théorique</i>	23
<i>I.1. Traitement de la base de données</i>	23
<i>I.2. Codification des modalités</i> :	23
<i>II. Cadre applicatif</i>	24
<i>II.1. Définition des variables</i> :.....	24
<i>II.1.1. Variable cible</i> :	24
<i>II.1.2 Variables qualitatives</i>	24
<i>II.1.3. Variables quantitatives</i>	26
<i>II.1.4. Variables Croisées</i> :.....	28
<i>II.2. Etude descriptive et agrégation avec le défaut</i>	28
<i>II.2.1. Variables qualitatives</i> :.....	28

II.2.2. Variables quantitatives :	33
II.2.3. Variables croisées :	35
Conclusion.....	37
Chapitre 2 Elaboration du modèle de score	38
I. Cadre théorique de l'élaboration d'un modèle de scoring.	39
I.1. Echantillonnage.	39
I.2. Corrélation entre les variables :	39
I.3. Méthode des réseaux de neurones :	39
I.3.1. Définition :	39
I.4. Régression logistique :	41
II. Elaboration du modèle de score.....	42
II.1. Echantillonnage :	42
II.2. Corrélation entre les variables :	43
II.3. Construction du modèle par la méthode des réseaux de neurones.....	43
II.3.1. Standarisation des données :	43
II.3.2. Approche PMC :	44
II.4. Régression logistique :	44
II.4.1. Défaut :	44
II.4.2. Méthode pas à pas descendante :	44
II.5. Contribution des variables retenues :	45
Conclusion.....	47
Chapitre 3: Validation du modèle.....	48
I. Cadre théorique.....	49
I.1. Taux de bon classement :	49
I.2. Courbe de ROC :	49
I.3. Indice de Gini :	50
I.5. Valeur de l'information :	51
II. Cadre applicatif.....	52
II.1. Méthode des réseaux de neurones :	52
II.1.1. Taux de bon classement :	52
II.1.2. Indice de GINI :	53
II.2. Méthode de la régression logistique :	54
II.2.1. Taux de bon classement :	54
II.2.2. Indice de GINI :	56
II.3. Valeur de l'information :	57
Conclusion :	58

Chapitre 4 : Elaboration de la grille de score	59
I. Méthodologie et élaboration de la grille	60
I.1. Calcul du score :	60
I.2. Grille de score :	61
II Tests sur la grille de score.....	64
II.1. Indice de performance (IP) :	64
II.1.1. Stabilité de la distribution des modalités de chaque variable.	64
II.1.2. Indice de performance (IP).....	69
II.1.3. Test de stabilité de la grille : Kolmogorov-Smirnov.....	71
Chapitre 5 : Automatisation du modèle de score	73
I. Automatisation	74
II. Sauvegarde des informations obtenues.	76
CONCLUSION GENERALE.	77
ANNEXES.....	79
Annexe I : Présentation de l'organisme	80
Annexe II: Présentation du logiciel utilisé.....	86
Annexe III : Matrice de corrélation des variables	84
Annexe IV : Calcul de la valeur de l'information de quelques variables.....	85
Annexe V : Réseau de neurones PMC (produit).....	86
Annexe VI : Sauvegarde des résultats (VBA Excel)	87

Liste des abréviations

CA : Crédit Automobile Classique

HCP: Haut Commissariat au plan

LOA : Location avec option d'achat

PP: Prêt personnel

ROC: Receiver Operating Characteristic

SPSS: Statistical package for the social science

TPE : Très petite Entreprise

Liste des tableaux

Tableau 1 : Codification de la variable « type-société ».....	23
Tableau 2 : Codification de la variable « secteur d'activité ».....	23
Tableau 3 : Codification de la variable « région ».....	23
Tableau 4 : Codification de la variable « catégorie produit ».....	24
Tableau 5 : Codification de la variable « location ».....	24
Tableau 6 : Codification de la variable « gamme véhicule ».....	24
Tableau 7 : Codification de la variable « durée de crédit ».....	25
Tableau 8 : Codification de la variable « montant capital ».....	25
Tableau 9 : Codification de la variable « apport ».....	25
Tableau 10 : Codification de la variable « ancienneté ».....	26
Tableau 11 : Taux de défaillance des modalités de la variable « type-société ».....	26
Tableau 12 : Codification de la variable « région »	27
Tableau 13 : Taux de défaillance des modalités de la variable « Région ».....	28
Tableau 14 : Taux de défaillance des modalités de la variable « location ».....	28
Tableau 15 : Codification de la variable « secteur d'activité ».....	29
Tableau 16 : Taux de défaillance des modalités de la variable « secteur d'activité ».....	29
Tableau 17 : Taux de défaillance des modalités de la variable « gamme véhicule ».....	30
Tableau 18 : taux de défaillance des modalités de la variable « catégorie produit ».....	31
Tableau 19 : taux de défaillance des modalités de la variable « ancienneté ».....	31
Tableau 20 : Statistiques descriptives de la variable « Apport ».....	31
Tableau 21 : taux de défaillance des modalités de la variable « Apport ».....	32
Tableau 22 : taux de défaillance des modalités de la variable « Montant Capital ».....	32
Tableau 23 : Statistiques descriptive de la variable durée de « crédit ».....	32
Tableau 24 : taux de défaillance des modalités de la variable « Durée de crédit ».....	32
Tableau 25 : taux de défaillance des modalités de la variable croisée « Secteur*Ancienneté ».....	33
Tableau 26 : Agrégation des modalités de la variable croisée « Ancienneté*Secteur ».....	33
Tableau 27 : taux de défaillance des modalités de la variable croisée « Location*Apport ».....	34
Tableau 28 : Agrégation des modalités de la variable croisée « Location*Apport ».....	34
Tableau 29 : Modèle de régression Logistique a l'étape 3.....	43
Tableau 30: Contributions des variables dans le modèle de la régression logistique.....	44
Tableau 31 : Matrice de confusion en effectif et en pourcentage pour la variable défaut « échantillon d'apprentissage ».....	51
Tableau 32 : Matrice de confusion en effectif et en pourcentage pour la variable défaut « échantillon test ».....	52
Tableau 33 : Contributions et valeur de l'information des variables du modèle.....	54
Tableau 34 : Système de notation.....	60

<i>Tableau 35 : Proportions des modalités de la variable « Code gamme » dans chaque échantillon.....</i>	<i>64</i>
<i>Tableau 36 : Proportions des modalités de la variable « Code secteur » dans chaque échantillon.....</i>	<i>64</i>
<i>Tableau 37 : Proportions des modalités de la variable « Catégorie produit » dans chaque échantillon.....</i>	<i>65</i>
<i>Tableau 38 : Proportions des modalités de la variable « code région » dans chaque échantillon.....</i>	<i>65</i>
<i>Tableau 39 : Proportions des modalités de la variable « Ancienneté » dans chaque échantillon.....</i>	<i>66</i>
<i>Tableau 40 : Proportions des modalités de la variable « Mnt Capital » dans chaque échantillon.....</i>	<i>66</i>
<i>Tableau 41 : Proportions des modalités de la variable « Apport » dans chaque échantillon.....</i>	<i>67</i>
<i>Tableau 42 : Proportions des modalités de la variable « Durée de crédit» dans chaque échantillon.....</i>	<i>68</i>
<i>Tableau 43 : Test de Kolmogorov Smirnov pour les non défaillants.....</i>	<i>70</i>
<i>Tableau 44 : Test de Kolmogorov Smirnov pour l'ensemble de la population.....</i>	<i>70</i>

Liste des figures

Figure 1: Proportion de chaque modalité de la variable <i>type_société</i>	28
Figure 2: Proportion de chaque modalité de la variable « région »	29
Figure 3: Proportion de chaque modalité de la variable « location »	30
Figure 4: Proportion de chaque modalité de la variable <i>secteur d'activité</i>	31
Figure 5: Proportion de chaque modalité de la variable « gamme véhicule »	32
Figure 6: Proportion de chaque modalité de la variable « catégorie-produit »	32
Figure 7: Le taux de défaut dans la base de données.	44
Figure 8: Contributions des variables dans le modèle de la régression logistique	46
Figure 9: Courbe de ROC	50
Figure 10: Courbe de Gini	51
Figure 11: proportions des modalités de la variable « Code gamme » dans chaque échantillon	64
Figure 12: proportions des modalités de la variable « Code secteur » dans chaque échantillon	65
Figure 13: proportions des modalités de la variable « catégorie produit » dans chaque échantillon ..	66
Figure 14: Proportions des modalités de la variable « Code région » dans chaque échantillon	66
Figure 15: Proportions des modalités de la variable « Ancienneté » dans chaque échantillon	67
Figure 16: proportions des modalités de la variable « Mnt capital » dans chaque échantillon	68
Figure 17: proportions des modalités de la variable « Apport » dans chaque échantillon	68
Figure 18: proportions des modalités de la variable « Durée de crédit » dans chaque échantillon ...	69

Introduction générale

Dans un contexte de crise, le monde se mobilise pour assurer une bonne gestion de risque.

Le comité de Bâle, dans cette perspective a édicté une multitude de règles prudentielles au sein des banques et des institutions financières dans le but de quantifier le risque, notamment le risque de crédit qui fait l'objet de notre étude. Parmi les mesures prudentielles édictées par Bâle II, l'établissement d'un système de notation interne au sein des banques et des établissements financiers.

Wafasalaf, étant un leader du crédit à la consommation au Maroc se doit de mesurer, afin de maîtriser le degré risque d'un client. Ainsi, toutes les unités de Wafasalaf se mobilisent auprès du pôle risque pour assurer une meilleure classification des clients en fonction de leur profil risque.

Et conformément aux normes de Bâle II, Wafasalaf a mis en place un système de notation interne pour les prêts personnels (PP), mais ne dispose pas actuellement un système de notation pour les sociétés. Notre projet de fin d'étude a pour objectif d'assurer cette notation interne des sociétés par l'élaboration d'une grille de score pour les Très Petites Entreprises (TPE) et ainsi permettre aux agents de se baser sur une règle de décision quant à l'octroi ou le refus d'un dossier de crédit.

Cependant, le travail effectué constitue pour l'organisme une première aide à la décision en attendant la collecte d'informations plus pertinentes et plus décisives comme il sera expliqué dans le corps du rapport.

Les deux premières parties du rapport comportent les étapes relatives à la modélisation du défaut, à commencer par le traitement de la base de données qui constitue l'étape clé de notre étude. La partie suivante a pour objectif de valider le modèle théorique pour ainsi construire, dans la partie 4 la grille de score. La dernière partie constitue une automatisation du modèle de score sous VBA Excel.

Chapitre 0 :
Généralités et présentation du
cadre de travail

I. Définitions et Généralités

I.1. Introduction

Toute institution financière, dans l'exercice de ses fonctions est sujette à des risques financiers. Ces risques financiers nécessitent un mécanisme de gestion interne.

On en distingue six principaux types :

- Le risque de liquidité : découle d'une difficulté à mobiliser des fonds à un coût raisonnable ou de la difficulté de vendre des actifs financiers à des prix proches de ceux du marché.
- Le risque de marché concerne les activités bancaires et les activités commerciales des banques, c'est un risque de perte en cas de mouvements défavorables des taux d'intérêts, des prix des matières premières et des cours des devises.
Les risques de marché peuvent être classés en risques liés aux prix des actions, risque de change, risque lié au taux d'intérêt ...
- Le risque de taux : concerne la variation de la valeur des fonds propres ou de la rentabilité suite à une variation du taux d'intérêt.
- Le risque de solvabilité : concerne l'insuffisance en réserves pour absorber les pertes éventuelles.
- Le risque de crédit : est le plus important et celui auquel sont confrontés tous les établissements de crédit de nos jours, il peut être géré s'il est bien mesuré d'où l'intérêt de développer un mécanisme de gestion interne de ce type de risque.
Dans ce qui suit, on va nous intéresser uniquement au risque de crédit.

I.2. Définition risque de crédit

Le risque de crédit se traduit par l'incapacité ou le refus d'un emprunteur de remplir ses obligations stipulées dans le contrat établi avec l'établissement de crédit.

En terme probabiliste le risque de crédit représente la probabilité de défaut de paiement d'un client, mais ne se limite pas uniquement à ce risque de défaillance mais aussi à d'autres risques qui peuvent compromettre la rentabilité de l'établissement financier.

Les plus importants sont :

- Le risque de portefeuille : est relatif à la dégradation du portefeuille.
- Le risque de concentration : survient suite à l'exposition sur un secteur particulier ou sur une zone donnée.
- Le risque pays : dû à une instabilité du pays en question qui pourrait engendrer une dégradation de sa note souveraine et donc de sa qualité de remboursement.

Dans cette perspective, les établissements financiers se doivent d'assurer une réserve en fonds propres dans le cadre de mesures prudentielles en respectant les proportions édictées par la réglementation de Bale, et ceci en vue de faire face aux pertes engendrées par la réalisation des risques auxquels ils sont confrontés.

I.3. Ratio standards :

Il n'y a encore que quelques années, l'octroi du crédit par une institution financière était une opération très classique. Les agents se contentaient d'étudier les rapports comptables de ceux qui sollicitaient des fonds en comparant leurs ratios financiers à ceux qui sont associés aux normes de bonne santé financière.

Avant de conclure la demande de crédit d'une entreprise, la banque se doit d'établir un diagnostic financier de l'entreprise qui demande le crédit en examinant son bilan (renseigne sur le patrimoine) et son compte de produits et de charges (donne le résultat de l'entreprise).

Cette analyse du risque a pour objectif d'étudier les comptes financiers passés de l'entreprise en vue de diagnostiquer ceux du présent et prévoir la santé financière de l'entreprise dans le futur.

Le diagnostic financier repose sur des ratios financiers représentés dans 3 grandes catégories :

✓ Ratios de structure :

Intitulé	Formule
Ratio d'autonomie financière	Capitaux propres/ capitaux permanents
Ratio de trésorerie immédiate	Disponibilité/ dettes à court terme
Ratio d'équilibre financier	Capitaux permanents/ actif immobilisé

✓ Ratios d'activité :

Intitulé	Formule
Part des frais financiers dans la valeur ajoutée	Charges financières//valeur ajoutée
Ratio crédit fournisseurs en mois	(Dettes fournisseurs/ achats TTC+ autres charges externes TTT)*12
Ratio crédit clients en mois	(Créances clients/CA TTC)*12

✓ Ratios de rentabilité :

Intitulé	Formule
Rentabilité financière	Résultat net de l'exercice/ capitaux propres

Ces ratios sont classés dans une catégorie de risque qui permet de fixer la prime de risque sur une base de jugement.

Aujourd'hui, des modèles sont développés pour modéliser le risque de crédit d'une institution financière, dont le modèle Scoring qui fera l'objet de notre étude.

I.4. Définition du crédit Scoring :

Le « Credit Scoring » est le processus d'attribution d'une note (ou score) à un emprunteur potentiel pour estimer la performance future de son prêt (Flaman, 1997).

En vue d'octroyer un crédit à un client, un établissement de crédit tout comme la banque a besoin de s'assurer que le client est solvable, c'est-à-dire qu'il est en capacité de rembourser ses emprunts. La banque exige donc des garanties à l'emprunteur, ces garanties ne suffisent pas toujours à trancher quant à la décision d'octroyer le crédit. Les établissements financiers ont donc recours à un système de score qui consiste à évaluer l'emprunteur en se basant sur plusieurs facteurs discriminants notamment le revenu, l'âge, le secteur d'activité ... et permettre de déterminer ainsi le profil risque du demandeur.

Le « Credit Scoring » n'approuve ni ne rejette une demande de prêt, il prédit plutôt la probabilité d'occurrence d'une mauvaise performance.

Il est plus utilisé par les établissements de crédit que par les banques, car celles-ci détiennent des dossiers en interne de leurs clients et ont donc plus d'informations concernant la qualité d'emprunteur que les établissements de crédit.

I.5. Avantages et inconvénients du « Credit Scoring » :

Pour faciliter la gestion du risque de crédit, les établissements financiers se doivent de développer un système qui retrace la qualité d'un engagement.

L'instauration d'une grille de notation interne s'avère donc essentielle et présente de nombreux avantages :

Le Credit Scoring est un procédé simple, qui permet d'améliorer la productivité des analyses de demande de crédit en traitant les dossiers et ainsi trancher rapidement quant à l'octroi ou non du crédit en se basant sur le score obtenu à partir des informations de l'emprunteur, ce qui engendre une satisfaction des clients qui voient leur demande traitée dans un temps très court.

Le Credit Scoring a aussi comme avantage d'être homogène, c'est-à-dire que les décisions issues de différentes agences sont les mêmes vu qu'elles s'appuient sur la même grille de notation construite à partir de critères qui se veulent objectifs.

Il peut être testé avant utilisation, et mesure aussi la probabilité de défaut de remboursement, il donne des tarifications de crédit adaptées au risque encouru et assure une meilleure allocation des fonds propres sous les directives de Bâle II.

Par ailleurs, l'utilisation de score assure à la banque une baisse de taux des impayés tout en conservant une clientèle avec un risque moyen à faible et dont le taux d'intérêt est déterminé selon la catégorie d'appartenance du risque.

Cependant, le « Credit Scoring » représente l'inconvénient d'appuyer la décision d'acceptation ou de rejet du dossier de crédit uniquement sur des chiffres, sans permettre au client de défendre son projet. C'est pour ceci que le Credit Scoring concerne généralement les petits crédits tels que les crédits à la consommation qui engendrent des petits montants et la prise en considération du facteur temps.

I.6. Cas de Wafasalaf :

Étant un établissement de crédit, la modélisation du risque de crédit s'avère être indispensable à Wafasalaf.

C'est dans cette perspective que nous sommes amenés à élaborer un modèle de scoring pour l'octroi de crédit adapté aux très petites entreprises (TPE) et quelques petites et moyennes entreprises (PME).

En général, les variables discriminantes sont prises à partir des ratios du bilan et du compte des produits et charges de l'entreprise.

Or Wafasalaf est un leader national dans le secteur du crédit et se base donc sur des crédits de masse. Ainsi le choix de telles variables discriminantes ne serait pas judicieux dans la mesure où l'étude de chaque dossier de crédit sera longue, d'où le besoin de déterminer d'autres variables qui pourraient être discriminantes tel que l'âge, la région....

I.7. Secteur TPE au Maroc :

La TPE est une entreprise dont le chiffre d'affaires ne dépasse pas les 3 millions de DH et qui n'emploie pas plus de 10 salariés.

Elle constitue la principale unité économique du pays, Selon les estimations avancées par le gouvernement, la Direction Générale des Impôts, l'Ompic et le HCP, elles constituent 3 millions d'unités (dont seulement 30.000 inscrites à la CNSS le secteur emploi a lui seul a peu près 6 million de personnes c'est-à-dire plus de 75% de la main d'œuvre du pays, il représente donc la principale source d'emploi.

Selon le Haut commissariat au plan (HCP) à peu près 80% des entreprises marocaines sont considérées comme étant des TPE.

Par ailleurs, les TPE contribuent pour 46% des opportunités d'emploi créées chaque année. De plus, il se crée chaque année 40.000 nouvelles TPE.

Devant une telle force économique, le secteur bancaire commence à s'intéresser de plus en plus à ce secteur et ne se restreint plus aux grandes entreprises seulement; et c'est dans cet

esprit que Wafasalaf a décidé de s'ouvrir aux TPE, et son activité ne se trouve plus réduite au crédit de consommation pour les particuliers.

II. Présentation du cadre du travail

II.1. Préparation des données :

En vue d'effectuer un Scoring de crédit pour les très petites entreprises, le département de prévention risque de Wafasalaf nous a fourni une base de données de 8540 observations qui regroupe les informations sur les tiers notamment leur type (SA, SARL ...), leur secteur d'activité..., et les informations concernant le crédit tels que la durée de crédit, le montant accordé, l'apport, la mensualité ...

Nous avons aussi jugé très intéressant d'obtenir des informations comme les ratios financiers tirés du Bilan et du CPC. Cependant, ces informations ne sont pas disponibles dans la base de données clients de Wafasalaf ce qui nous a contraint à nous restreindre aux informations disponibles uniquement.

I. II.1.1. Critère et variable de défaut :

Le comité de Bâle II définit le défaut de la façon suivante:

« Un défaut de la part d'un débiteur intervient lorsque l'un des événements ci-dessous se produit:

- La banque estime improbable que le débiteur rembourse en totalité son crédit au groupe bancaire sans qu'elle n'ait besoin de prendre des mesures appropriées telles que la réalisation d'une garantie (si elle existe).

-L'arriéré du débiteur sur un crédit important dû au groupe bancaire dépasse 90 jours »

Dans notre étude, on définit le critère de défaut conformément aux normes de Bâle II :

- ✓ Critère de défaut : Cumul de 3 impayés (90 Jrs de non remboursement)
- ✓ Variable de défaut : variable définissant le critère de défaut

$$\text{DEFAULT} = \begin{cases} 1 & \text{si l'entreprise est défailante} \\ 0 & \text{sinon} \end{cases}$$

II.1.2. Fenêtre d'observation :

La fenêtre d'observation comporte l'ensemble des observations contribuant dans la construction du modèle.

Notre base de données comporte une fenêtre d'observation de 3ans (2008 à 2011).

Après l'élaboration du modèle de Score, il convient d'exploiter ses résultats en les appliquant sur un échantillon récent pour confirmer la stabilité dans le temps.

La durée que couvre notre échantillon récent est dite fenêtre de validation et s'étale sur 6 mois à compter du mois d'Octobre 2012.



Chapitre 1 : ***Données et méthodologie de l'étude.***

Dans cette partie nous allons procéder au nettoyage et à la préparation de la base de données.

Puis, une étude statistique sera réalisée pour chaque variable dans le but de mieux cerner le comportement de chacune et son degré d'implication dans le modèle.

I. Cadre théorique

I.1. Traitement de la base de données.

Avant d'entamer la modélisation, il est important de disposer d'une base de données complète et cohérente d'où la nécessité de traiter les valeurs manquantes et les valeurs aberrantes.

I.1.1. Traitement des valeurs manquantes :

Le traitement des variables manquantes se fait par le biais de la méthode d'imputation qui consiste à remplacer cette valeur manquante.

Il existe diverses méthodes d'imputation :

- Imputation par moyenne :

La valeur manquante est remplacée par la moyenne de l'ensemble des réponses.

- Imputation par ratio :

Consiste à remplacer la variable manquante Y_i par $Y_i = a * X_i$ tel $a = \bar{Y}_r / \bar{X}_r$

et \bar{Y}_r (resp \bar{X}_r) est la moyenne des réponses obtenues de la variable Y_i (resp X_i)

- Imputation par régression :

C'est une extension de la méthode de ratio qui consiste à utiliser plusieurs variables auxiliaires à la place d'une seule.

- Imputation par plus proche voisin.

On cherche parmi les individus ayant répondu aux questions celui dont les réponses aux autres questions sont le plus proches des réponses de l'individu dont la valeur est manquante. Il s'agit d'établir une fonction distance à partir de laquelle on déduit les non réponses.

I.1.2. Traitement des valeurs aberrantes :

Les valeurs aberrantes sont des valeurs dont l'observation ne semble pas suivre la tendance générale et dévie du comportement global des autres variables.

Pour les repérer, plusieurs méthodes peuvent être exploitées notamment le diagramme de dispersion, la boîte à moustache ou une étude descriptive des variables.

Une fois repéré, il est nécessaire de les traiter.

Ce traitement se fait par une correction de la valeur si évidente ($\hat{\text{âge}} < 0$), l'utilisation des techniques d'imputation comme pour les valeurs manquantes ou encore les éliminer définitivement.

I.2. Codification des modalités :

Dans un souci de simplification et pour éviter une écriture encombrée, on procède à une codification des modalités de chaque variable avant de commencer la modélisation par la régression logistique.

II. Cadre applicatif

II.1. Définition des variables :

II.1.1.. Variable cible :

La variable à expliquer est la défaillance du client, elle définit le critère de défaut. C'est une variable binaire notée :

$$\triangleright \text{Défaut} = \begin{cases} 1 & ; \text{Si l'entreprise est défaillante.} \\ 0 & ; \text{Sinon.} \end{cases}$$

En ce qui concerne les variables explicatives retenues pour la construction du modèle, on distingue trois types de variables :

- Variables à caractère quantitative.
- Variables à caractère qualitative.
- Variables Croisées.

II.1.2 Variables qualitatives

Les variables qualitatives disponibles pour l'élaboration de notre modèle sont les suivantes :

✓ Type-société :

Distingue entre les différents types de sociétés existantes et en relation avec Wafasalaf. Et qui sont regroupées comme suit :

On distingue :

- SA : société anonyme, comprend les sociétés anonymes à directoires et les sociétés anonymes à conseil d'administration.
- SARL : Société anonyme à responsabilité limitée
- SARL GU : Société anonyme à responsabilité limitée, à gérant unique.
- Autres : regroupe les sociétés en nom collectif, les sociétés civiles, les sociétés de fait et les sociétés personnelles.

<u>CODE</u>	<u>Libellé</u>
1	SARL
2	SARL GU

Tableau 1 : Codification de la variable type-société

✓ Grand secteur : définit les secteurs d'activité des entreprises.

Notre base de données initiale comprend la variable « SEC_ACTIVITE » qui regroupe tous les secteurs d'activité. Cependant, vu le nombre très important de modalités de cette variable, nous avons décidé de l'agréger en créant une nouvelle variable que l'on a appelée « GRAND_SECTEUR ». On distingue ainsi :

<u>CODE</u>	<u>Libellé</u>
1	BTP
2	COMMERCE DETAIL
3	INDUSTRIE
4	LOCATION
5	SERVICES
6	TRANSPORT

Tableau 2 : Codification de la variable secteur d'activité

✓ Région : indique la région où opère l'entreprise.

La base de données comprend un très grand nombre de clients ayant leur siège dans diverses régions du Maroc que l'on a pu rassembler selon les régions suivantes auxquelles on a attribué un code :

<u>Code</u>	<u>Libellé</u>
1	AGADIR_SUD
2	CASA_SETTAT
3	EIJADIDA_DOUKALA
4	FES_SAISS
5	KENITRA_GHARB
6	MARRAKECK_HAOUZ
7	MEKNES_CENTRE
8	OUJDA_ORIENT
9	RABAT_SALE
10	TANGER_NORD
11	AUTRES

Tableau 3 : Codification de la variable région

✓ Catégorie-Produit :

Cette Variable nous permet de distinguer entre les deux types de produits faisant l'objet d'un prêt, à savoir le crédit automobile noté CA et la location avec Option d'Achat :

<u>CODE</u>	<u>Libellé</u>
1	LOA
2	CA

Tableau 4 : Codification de la variable catégorie-produit

Le Crédit Auto (CA) est un prêt amortissable destiné aux personnes physiques et morales pour le financement de leurs projets d'acquisition de véhicule neufs.

La location avec Option d'Achat (LOA) est une formule de financement destinée à l'acquisition de biens d'équipement tel que les véhicules neufs avec exonération faite du montant de la TVA. Ceci permet au client de jouir du bien sans en être propriétaire, avec possibilité de l'acquérir en fin de contrat.

✓ Loueur/Hloueur :

Renseigne selon si l'entreprise est une entreprise de location, ou opère dans d'autres secteurs d'activité.

Le tableau ci-dessous nous donne la codification choisie pour cette variable.

<u>CODE</u>	<u>Libellé</u>
1	hors location
2	Location

Tableau5 : Codification de la variable location

✓ Gamme-Voiture : définit la gamme de la voiture pour laquelle le crédit a été octroyé.

Pour cette variable, nous avons opté pour la classification suivante:

<u>CODE</u>	<u>Libellé</u>
1	HAUT de GAMME
2	MOYENNE GAMME
3	LOW-COST

Tableau 6 : Codification de la variable gamme véhicule

II.1.3. Variables quantitatives

On a procédé a une discrétisation de ces variables .Ainsi chaque variable quantitative a été découpée en plusieurs intervalles que l'on a codifié par la suite.

Le choix des classes de chacune de ces variables ne s'est pas fait de manière anodine. Le sens du risque a été pris en considération.

Ainsi les variables quantitatives choisies sont :

- ✓ DUREE_CRED: indique la durée du prêt en mois.

<u>Code</u>	<u>Classes</u>
1	moins de 36 mois
2	[36 mois;60 mois [
3	plus de 60 mois

Tableau 7 : Codification de la variable durée crédit

- ✓ MNT_CAPITAL: représente le montant du capital de l'entreprise.

<u>Code</u>	<u>classes</u>
1	moins de 100 000 DHS
2	Plus de 100 000 DHS

Tableau 8: Codification de la variable montant-capital

- ✓ Apport% :

Constitue le pourcentage apporté par le client et est défini par le rapport entre le montant apporté et le montant du véhicule.

<u>Code</u>	<u>Classes</u>
1	moins de 37%
2	[37%; 48% [
3	plus de 48%

Tableau 9 : Codification de la variable apport

- ✓ Ancienneté de l'activité :

Définit l'ancienneté de l'entreprise en mois et est définie par la différence entre la date de début d'activité et la date de début de relation avec Wafasalaf.

<u>Code</u>	<u>Classes</u>
1	moins de 18 mois
2	[18 mois; 40 mois [
3	Plus de 40 mois

Tableau 10: Codification de la variable Ancienneté

II.1.4. Variables Croisées :

Les variables croisées sont des variables constituées à partir des variables initiales et qui donnent une information supplémentaire.

Parmi les variables que l'on a jugé bon de croiser

Apport*location
Ancienneté*secteur

II.2. Etude descriptive et agrégation avec le défaut.

II.2.1. Variables qualitatives :

NB: Il est à noter que les taux de risque de WAFASALAF pour chaque variable sont soumis à une contrainte de confidentialité.

Nous exposerons ci-dessous la démarche effectuée tout en masquant ces taux.

✓ Type société :

Les 4 types sont répartis selon les proportions suivantes :


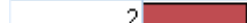


Valeur	Proportion	%	Comptage
1		82,44	6809
2		13,86	1145
3		1,19	98
4		2,51	207

Figure 1: Proportion de chaque modalité de la variable type société

On remarque que la classe 3 (resp la classe 4) représente 1.19% (resp 2.51%) de la population globale. Ces proportions sont très insuffisantes. Il faut donc redéfinir les classes.

Pour se faire, nous allons agréger la variable type de société avec le défaut pour ainsi rassembler les classes qui représentent le même risque de défaillance et obtenir des effectifs > 5%.

Après agrégation, il convient de rassembler les classes 2,3 et 4 dans une même classe et garder la classe 1.

On obtient :

<u>code-soc</u>	<u>défaut Sum</u>	<u>Effectif</u>	<u>%</u>
1	β_{11}	7015	α_{11}
2	β_{12}	1524	α_{12}

Tableau11 : taux de défaillance des modalités de la variable « Ancienneté »

Les taux de risque pour les deux classes sont presque égaux. La variable « type_société » n'est pas une variable discriminante.

Par la suite, et lors de la régression logistique, cette variable n'entrera sûrement pas dans le modèle.

✓ Région :

La base de données de Wafasalaf contient la région où opère chaque client, cette variable peut aussi être intéressante dans la modélisation dans le sens où certaines régions sont plus risquées que d'autres en termes de remboursement de prêt.

L'ensemble de ces régions est réparti comme suit :












Valeur	Proportion	%	Comptage
AGADIR-SUD		9,98	824
AUTRES		3,5	289
CASA-SETTAT		30,11	2487
ELJADIDA-DOUKA...		2,57	212
FES-SAISS		5,4	446
KENITRA-GHARB		3,64	301
MARRAKECH-HA...		12,24	1011
MEKNES-CENTRE		2,78	230
OIJDA-ORIENT		3,06	253
RABAT-SALE		16,08	1328
TANGER-NORD		10,63	878

Figure 2: Proportion de chaque modalité de la variable « région »

Les crédits octroyés par Wafasalaf se concentrent particulièrement dans les régions Casa-Settat, Rabat-Sale et Marrakech Haouz.

La région Meknès-centre est la moins représentée avec un pourcentage de 2.2 % seulement.

Il faudrait donc rassembler certaines régions en prenant en compte le risque de défaut afin d'obtenir des classes avec des effectifs >5%.

On obtient donc la classification suivante :

<u>code-région</u>	<u>Libellé</u>
1	Casa- Settat/ Meknès-centre
2	Kenitra-Gharb /Fès-Saïs/ Marrakech-Haouz / Rabat-Sale/ Agadir-Sud/ Autres
3	Tanger-Nord / Oujda-Orient /Jdida-Doukala

Tableau 12 : Codification de la variable « région »

Leur agrégation avec le défaut donne les pourcentages suivants :

<u>code-région</u>	<u>défaut Sum</u>	<u>Effectif</u>	<u>%</u>
1	β_{21}	2833	α_{21}
2	β_{22}	4313	α_{22}
3	β_{23}	1393	α_{23}

Tableau 13 : taux de défaillance des modalités de la variable « région »

On remarque donc que le risque varie d'une classe à l'autre pour passer de 14% pour la classe 1 à 22% pour la classe 3.

Cette variable peut être discriminante.

✓ Location / HLocation :

Les proportions des modalités de cette variable sont les suivantes :


Valeur	Proportion	%	Comptage
1		82,82	7072
2		17,18	1467

Figure 3:Proportion de chaque modalité de la variable « location »

On remarque donc que la majorité des sociétés clientes de Wafasalaf sont des sociétés de type hors location et représentent environs 82 %, tandis que les sociétés de location représentent environ 18%.

L'agrégation de cette variable avec la variable défaut nous donne:

<u>code-location</u>	<u>défaut Sum</u>	<u>Effectif</u>	<u>%</u>
1	β_{31}	7072	α_{31}
2	β_{32}	1467	α_{32}

Tableau 14 : taux de défaillance des modalités de la variable « location »

Les sociétés de location et d'hors location présentent des probabilités de défaillance très proches. Cette variable est donc à première vue non discriminante, ce qui nous laisse penser qu'elle ne sera pas prise par notre modèle.

✓ Secteur d'activité

Le tableau ci dessous représente la proportion de chaque secteur d'activité dans notre base de données.

Valeur	Proportion	%	Comptage
BTP		14,16	637
COMMERCES_DET...		11,07	498
INDUSTRIE		6,53	294
LOCATION		24,54	1104
SERVICES		42,14	1896
TRANSPORT		1,56	70

Figure 4: Proportion de chaque modalité de la variable secteur d'activité

Certains effectifs sont très faibles, d'où la nécessité de rassembler quelques secteurs.

Après agrégation avec le défaut, on remarque que le secteur de l'industrie, transport et BTP représentent les secteurs les plus risqués avec un pourcentage de 20% environ. Les autres secteurs représentent un risque moyen d'environ 17%.

On obtient donc 2 classes distinctes :

<u>CODE</u>	<u>Libellé</u>
1	COMMERCE DETAIL/LOCATION/SERVICES
2	INDUSTRIE/TRANSPORT/BTP

Tableau 15 : « Codification de la variable secteur d'activité »

Le pourcentage de défaut de ces 2 classes varie de 8%. Cette variable est assez discriminante.

<u>code secteur act</u>	<u>défaut Sum</u>	<u>Effectif</u>	<u>% défaut</u>
1	β_{41}	6803	α_{41}
2	β_{42}	1736	α_{42}

Tableau16 : taux de défaillance des modalités de la variable « secteur d'activité »

✓ Gamme-voiture :

Ci dessous les proportions de chaque modalité de la variable Gamme-voiture dans notre base de données :




Valeur	Proportion	%	Comptage
1		15,99	1321
2		60,12	4965
3		23,89	1973

Figure 5:Proportion de chaque modalité de la variable « gamme véhicule »

Pour cette variable nous remarquons que la majorité des TPE (environ 60%) prennent des crédits de Wafasalaf pour l'achat de véhicules de classe 2 c'est-à-dire de moyenne gamme, loin devant la classe des low-cost ou celle de la moyenne gamme qui représentent respectivement une part de marché d'environ 24% et 16%.

Ensuite nous avons procédé à l'agrégation de cette variable codifiée avec le défaut en vue de dégager le taux de défaillance de chacune de ces classes.

Nous obtenons donc un taux de défaillance pour la moyenne gamme supérieur au taux de défaillance pour les véhicules haut de gamme et les véhicules low cost. Ces dernières ont un taux de défaillance quasi-similaire, on a donc jugé bon de les rassembler en 1 seule classe (classe 1), ce qui nous permet d'obtenir le tableau suivant :

<u>code gamme</u>	<u>défaut Sum</u>	<u>Effectif</u>	<u>%</u>
1	β_{51}	3389	α_{51}
2	β_{52}	5150	α_{52}

Tableau17 : taux de défaillance des modalités de la variable « gamme véhicule ».

La variable Gamme peut donc être introduite dans le modèle.

✓ Catégorie produit :

Le tableau ci-dessous montre que la majorité des crédits accordés par Wafasalaf sont de type Crédit Auto contrairement au crédit leasing des sociétés, qui reste un produit nouveau avec une très faible part du marché.


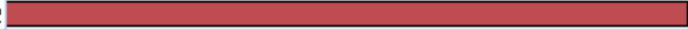
Valeur	Proportion	%	Comptage
1		6,56	560
2		93,44	7979

Figure 6:Proportion de chaque modalité de la variable « catégorie-produit »

L'agrégation de cette variable avec le défaut donne un pourcentage de défaut élevé pour le CA (18%) et un pourcentage de défaut significativement faible (4%), ceci est dû au faible effectif de la modalité LOA mais aussi au fait qu'un crédit leasing est par définition moins risqué qu'un crédit classique.

<u>categorie prod code</u>	<u>défaut Sum</u>	<u>effectif</u>	<u>%</u>
1	β_{61}	560	α_{61}
2	β_{62}	7979	α_{62}

Tableau 18 : taux de défaillance des modalités de la variable « catégorie-produit »

II.2.2. Variables quantitatives :

✓ Ancienneté :

La base de données de Wafasalaf comprend quelques sociétés dont l'ancienneté est négative. On y a remédié en remplaçant toutes les valeurs négatives par 0.

L'ancienneté d'une entreprise est discriminante. En effet, le taux de défaut est grand (23%) pour les nouvelles entreprises, et baisse pour atteindre les 12% pour les entreprises ayant une ancienneté de plus de 40 mois.

<u>Classes</u>	<u>Ancienneté</u>	<u>défaut Sum</u>	<u>Effectif</u>	<u>%</u>
moins de 18 mois	1	β_{71}	2877	α_{71}
[18 mois; 40 mois [2	β_{72}	2170	α_{72}
Plus de 40 mois	3	β_{73}	3490	α_{73}

Tableau 19 : taux de défaillance des modalités de la variable « Ancienneté »

✓ Apport :

D'après l'étude descriptive faite sur cette variable on remarque que l'apport en pourcentage des clients de Wafasalaf représente en moyenne 34%, et que la moitié des entreprises ont un apport supérieur ou égale à 33%.

<u>Comptage</u>	8537
<u>Moyenne</u>	34,85
<u>Min</u>	0
<u>Max</u>	94,856
<u>Variance</u>	220,615
<u>Médiane</u>	33,33

Tableau 20 : Statistiques descriptives de la variable « Apport »

L'agrégation de cette variable avec le défaut donne que le risque de défaut baisse avec l'augmentation de l'apport. Ce qui était prévisible.

<u>classes</u>	<u>Apport</u>	<u>défaut Sum</u>	<u>Effectif</u>	<u>%</u>
moins de 37%	1	β_{81}	4884	α_{81}
[37%; 48%[2	β_{82}	1935	α_{82}
plus de 48%	3	β_{83}	1718	α_{83}

Tableau 21 : Taux de défaillance des modalités de la variable « Apport »

✓ Montant capital :

Pour la variable montant de capital, le risque de défaut varie selon si le montant investi est supérieur ou inférieur à 100 000.

<u>Classes</u>	<u>MNT CAPITAL Tr</u>	<u>défaut Sum</u>	<u>Effectif</u>	<u>%</u>
moins de 100 000 DHS	1	β_{91}	2141	α_{91}
Plus de 100 000 DHS	2	β_{92}	6396	α_{92}

Tableau 22 : taux de défaillance des modalités de la variable « MNT Capital »

✓ Durée de crédit :

La durée de crédit dans l'échantillon varie de 1 mois à 77 mois, avec une moyenne de 40 mois.*

<u>Comptage</u>	8537
<u>Moyenne</u>	39,52
<u>Min</u>	1
<u>Max</u>	77
<u>Variance</u>	183,218
<u>Médiane</u>	40

Tableau 23 : Statistiques descriptives de la variable « Durée de crédit »

La durée de crédit a été découpée en 3 classes dont le risque est grand pour les courtes durées (<36 mois), puis diminue pour atteindre pour les durées moyennes et augmente encore pour les longues durées.

<u>Classes</u>	<u>DUREE CRED-TR</u>	<u>défait Sum</u>	<u>Effectif</u>	<u>%</u>
moins de 36 mois	1	$\beta_{10,1}$	2517	$\alpha_{10,1}$
[36 mois; 60 mois [2	$\beta_{10,2}$	4982	$\alpha_{10,2}$
Plus de 60 mois	3	$\beta_{10,3}$	1038	$\alpha_{10,3}$

Tableau 24 : Taux de défaillance des modalités de la variable « Durée de crédit »

II.2.3.Variables croisées :

En plus des variables quantitatives et qualitatives, il est aussi intéressant de croiser certaines variables en vue de tester leur pouvoir discriminant et ainsi l'introduire dans le modèle.

✓ Ancienneté * Secteur :

La première variable croisée que l'on a jugé intéressante est l'ancienneté * secteur dans le sens où cette variable nous donne une information sur chaque secteur selon son ancienneté.

Le tableau ci-dessous est un tableau croisé qui donne le taux de défaillance de chaque modalité de cette variable croisée.

		Ancienneté			
		% Défait	1	2	3
code_secteur_act	1		21,6	16,8	11,8
	2		27,6	21,4	12,1

*Tableau 25: taux de défaillance des modalités de la variable croisée « Secteur*Ancienneté »*

Pour une grande ancienneté, tous les secteurs d'activité ont la même probabilité de défaut.

Aussi, le risque diminue avec l'ancienneté et le secteur 1 est plus risqué que le secteur 2. C'est dans ce sens qu'on remarque que un secteur 2 nouveau est aussi risqué qu'un secteur 1 moyennement ancien. Il est bon de rassembler ces deux modalités.

On obtient donc la variable Ancienneté * Secteur ayant des valeurs G1, G2, G3 et G4.

		Ancienneté		
		1	2	3
code_secteur_act	% Défaut			
	1	G2	G3	
	2	G1	G2	G4

Tableau 26 : Agrégation des modalités de la variable croisée « Ancienneté*Secteur »

✓ Apport * location :

La deuxième variable à croiser est Apport * location qui renseigne sur l'apport selon si le client est loueur ou non loueur.

		APPORT%		
		1	2	3
Code-loueur	% défaut			
	1	20,20%	20,95%	$\alpha_{11,1}$
	2	20,16%	12,99%	$\alpha_{11,2}$

Tableau 27 : Taux de défaillance des modalités de la variable croisée « Location*Apport »

On remarque que certaines modalités ont une probabilité de défaut similaire, ce qui nous amène à les rassembler en une seule modalité. On obtient :

		APPORT%		
		1	2	3
Code-loueur	% défaut			
	1	G1		G2
	2	G2		G3

Tableau 28 : Agrégation des modalités de la variable croisée « Location*Apport »

✓ Autres variables croisées :

Nous avons aussi croisé d'autres variables qu'on a jugés cohérentes, cependant les résultats obtenus ne fournissent aucune information supplémentaire.

Conclusion

En guise de conclusion, on a constaté tout d'abord que notre base de données ne contenait aucun point aberrant.

Aussi nous avons remarqué d'après les statistiques descriptives menées, que tous les champs des variables qualitatives étaient bien renseignés, seules les variables quantitatives présentaient des valeurs manquantes que l'on a traitées grâce aux méthodes d'imputation.

Aussi en agrégeant chacune de nos variables avec le défaut réel, il s'est avéré nécessaire de rassembler plusieurs de ses modalités, afin d'obtenir des modalités significatives vis-à-vis du défaut

Chapitre 2 : Elaboration du modèle de score

Dans cette partie, nous allons entamer la construction du modèle via la méthode de la régression logistique afin de déterminer les variables discriminantes, leurs pouvoir discriminant, et estimer les paramètres du modèle.

I. Cadre théorique de l'élaboration d'un modèle de scoring.

I.1. Echantillonnage.

Après traitement de la base de données, il est nécessaire de la diviser en deux échantillons.

Le premier échantillon est l'échantillon d'apprentissage qui nous servira pour la construction du modèle et le second constitue l'échantillon test qui servira à valider le modèle.

Il est nécessaire de constituer les échantillons de façon à avoir la même probabilité de défaut que celle de la base de données, à savoir la même proportion de clients défaillants. Ces échantillons doivent être représentatifs de la population globale.

I.2. Corrélation entre les variables :

Les variables explicatives choisies ne devraient pas être corrélées entre elles.

La matrice de corrélation donne les coefficients de corrélation entre les variables et nous permet ainsi de distinguer les variables qui fournissent la même information et ainsi, d'éviter la redondance.

Deux variables explicatives ont une corrélation :

- Nulle si le coefficient de corrélation = 0
- Forte si le coefficient de corrélation est $\geq 0,7$ ou $\leq -0,7$.
- Parfaite si le coefficient de corrélation est égale -1 ou 1 (selon l'évolution dans le même sens ou dans le sens contraire).

I.3. Méthode des réseaux de neurones :

I.3.1. Définition :

Un réseau de neurones artificiel est un modèle de calcul dont la conception est très schématiquement inspirée du fonctionnement de vrais neurones.

Les réseaux de neurones sont des techniques sophistiquées de modélisation et de prévision, en mesure de modéliser des relations entre des données ou des fonctions particulièrement complexes. Ainsi ils ont une bonne faculté à donner un sens, extraire des règles et des tendances à partir de données compliquées.

Un réseau de neurones a une architecture calquée sur celle du cerveau, organisés en neurones et synapses, et se présente comme un ensemble de nœuds connectés entre eux, chaque variable continue en entrée correspond à un nœud d'un premier niveau, appelé couche d'entrée, et chaque modalité d'une variable qualitative correspond également à un nœud de la couche d'entrée. Lorsque le réseau est utilisé à des fins prédictifs, il y a une ou plusieurs variables à expliquer, elles correspondent chacune à un nœud dans la couche de sortie. Les réseaux prédictifs sont dits « à apprentissage supervisé » et les réseaux descriptifs sont dits « à apprentissage non supervisé ». Entre la couche d'entrée et la couche de sortie sont parfois

connectés des nœuds appartenant à un niveau intermédiaire : couche cachée. Enfin, un réseau de neurones est un ensemble de nœuds connectés entre eux, où chaque variable correspond à un nœud. Il y a plusieurs algorithmes dans les réseaux de neurones, dont les plus utilisées sont : Perceptron multicouches (PMC) et le Réseau à fonction radiale de base (RBF).

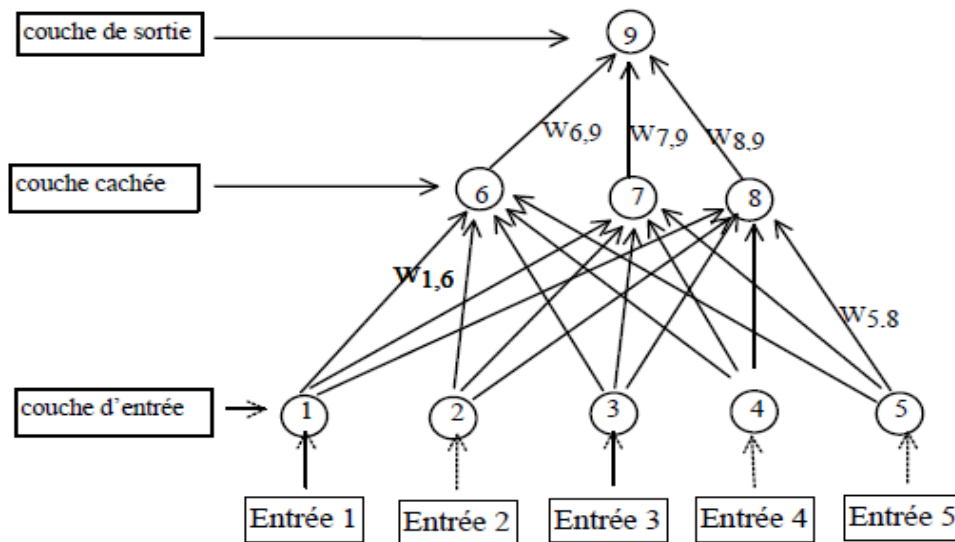


Figure : description d'un réseau de neurones

Les neurones d'un réseau possèdent des fonctions d'activation qui vont transformer les signaux émis par les neurones de la couche précédente à l'aide d'une fonction mathématique.

Il est à noter aussi que cette méthode suppose la normalisation des variables, et donc pour normaliser nos variables on peut utiliser la correspondance suivante :

$$\begin{aligned}
 0 &\rightarrow 0 \\
 1 &\rightarrow 1/2 \\
 2 &\rightarrow 1/2 + 1/4 \\
 &\vdots \\
 n &\rightarrow \sum_{k=1}^n 2^{-k}
 \end{aligned}$$

Le tableau ci-dessous décrit les avantages et les inconvénients de la modélisation via les réseaux de neurones :

Avantages	Inconvénients
<ul style="list-style-type: none"> • Ils prennent en compte les relations non linéaires et les interactions complexes entre les variables. • C'est une technique non paramétrique donc elle ne suppose pas d'hypothèses probabilistes sur les variables explicatives. • résiste plus aux données défectueuses. 	<ul style="list-style-type: none"> • L'impossibilité de traiter un grand nombre de variables. • Les résultats sont non explicites, ce qui est rédhibitoire pour certaines applications. • Nécessite la normalisation de chacune des variables.

Tableau : Avantages et inconvénients des réseaux de neurones.

I.4. Régression logistique :

Pour déterminer la probabilité pour qu'un client soit défaillant, on utilise la régression logistique via sa fonction « logit ».

I.4.1. Définition de la régression logistique:

La régression logistique est un modèle de régression multivariée, Elle s'utilise lorsque la variable à expliquer (variable dépendante Y) est qualitative, le plus souvent binaire. Les variables explicatives (variables indépendantes X_i) peuvent être qualitatives ou quantitatives.

La variable dépendante représente habituellement la survenance ou non d'un événement. Dans notre cas, si une entreprise est saine et ne représente aucun risque de non remboursement, la variable Y prendra la valeur 0 sinon la valeur 1.

Les variables indépendantes sont celles susceptibles d'influencer la survenue de cet événement, c'est-à-dire les variables mesurant l'exposition à un facteur de risque ou à un facteur protecteur, ou variable représentant un facteur de confusion.

La régression logistique s'avère être l'une des méthodes de modélisation les plus fiables, et dont plusieurs indicateurs statistiques permettent d'en contrôler facilement la robustesse (LR ratio, R carré de McFadden, Test de Hosmer-Lemeshow).

La régression logistique n'exige pas forcément une distribution normale des variables ni l'homogénéité des variances. Il est cependant important de disposer de grands échantillons.

La régression logistique permet de modéliser une probabilité directement.

L'inconvénient relatif à cette méthode est que l'analyse reste sensible aux individus ayant des valeurs aberrantes et aussi, ne peut être appliquée à une base de données avec des valeurs manquantes.

I.4.2. Formule :

La régression logistique est basée sur la formule suivante :

$$\ln\left(\frac{p}{1-p}\right) = \beta X = \beta_0 + \beta_1 X_1 + \dots + \beta_J X_J$$

Avec :

- ✓ (X_1, \dots, X_J) : les variables discriminantes.
- ✓ $\beta = (b_0, \dots, b_J)$: les coefficients de la fonction score.
- ✓ $P(1/X)$: représente la probabilité conditionnelle pour qu'un client soit défaillant.

Le principe consiste à égaliser la fonction « logit » à notre variable binomiale dépendante, pour déterminer la probabilité de non paiement d'un client.

Donc après transformation de notre hypothèse on obtient :

$$p_i = \frac{e^{\beta X_i}}{1 + e^{\beta X_i}}$$

- ✓ Estimation des paramètres β :

L'estimation se fait par la méthode de Maximum de vraisemblance dont la fonction est définie par :

$$l = \prod_{i=1}^N p_i^{y_i} (1 - p_i)^{1-y_i}$$

Il s'agit de maximiser cette fonction de vraisemblance. Ceci se fait par des méthodes numériques en utilisant des processus itératifs tel que la méthode de Newton Raphson.

- ✓ Interprétation des paramètres β :

Si $\beta_i > 0$, alors toute augmentation de la variable explicative X_i a un effet positif sur P_i .

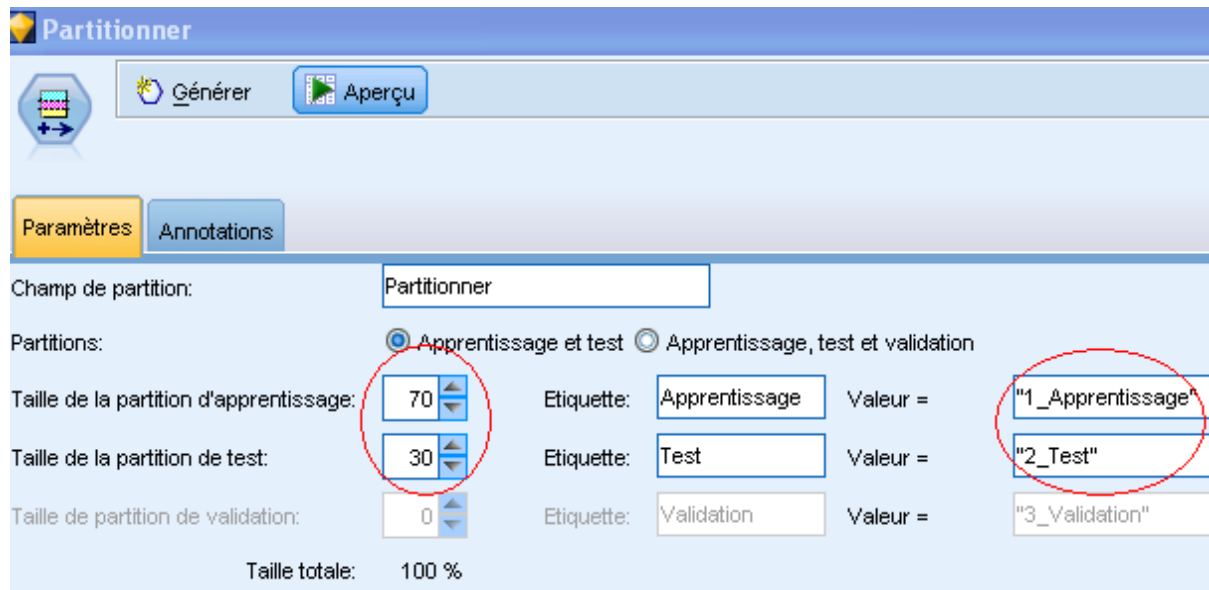
Si $\beta_i < 0$, alors toute augmentation de la variable explicative X_i a un effet négatif sur P_i .

I. Elaboration du modèle de score

II.1. Echantillonnage :

Après traitement de la base de données et codification des variables, on a procédé à la constitution d'échantillon d'apprentissage et de l'échantillon test.

La constitution des échantillons a été effectuée par SPSS Modeler à partir du nœud partitionner comme suit :



On obtient une nouvelle variable partitionner ayant pour valeurs « 1_apprentissage » et « 2_test ».

L'échantillon d'apprentissage constitue 70% de la l'échantillon global, c'est l'échantillon sur lequel on construit le modèle.

L'échantillon test représente 30% de l'échantillon global et sert à valider le modèle construit à partir de l'échantillon d'apprentissage et de tester sa stabilité et sa robustesse.

II.2. Corrélation entre les variables :

Avant d'entamer la régression logistique, il faut que nos variables soient non corrélées entre elles.

La matrice corrélation est présentée en annexe III.

On déduit à partir de la matrice de corrélation que la variable ancienneté*secteur est corrélée positivement avec la variable ancienneté.

Aussi, la variable apport*Loca est corrélée avec la variable apport.

Ceci dit, on choisit d'exclure la variable ancienneté*Secteur du modèle et de garder les variables ancienneté et secteur chacune à part vu qu'elles ne sont pas corrélés entre elles.

Pour ce qui est des variables apport et apport*loca, on choisit de tester le modèle avec chacune à part et retenir le modèle le plus robuste (courbe de GINI).

II.3. Construction du modèle par la méthode des réseaux de neurones.

II.3.1. Standarisation des données :

Avant d'entamer la construction du modèle, il faut s'assurer que les variables sont normées, comme il a été cité dans la partie théorique de ce même chapitre.

La standardisation a été effectuée sur toutes les variables suivant la méthode précédemment énoncée.

II.3.2. Approche PMC :

Le modèle des réseaux de neurones a été appliqué pour l'ensemble des variables explicatives.

L'approche PMC combine entre l'approche empirique et les caractéristiques des variables.

On obtient le réseau de neurone affiché en Annexe V.

Dans cette méthode, on retient 8 variables dans la couche d'entrée, on a aussi 5 neurones dans la couche cachée, et la variable de défaut dans la couche de sortie.

Par la suite, et lors de la validation du modèle. Cette méthode ne sera pas adoptée pour l'élaboration de la grille de score

II.4. Régression logistique :

II.4.1. Défaut :

Avant d'entamer la construction du modèle sur l'échantillon d'apprentissage, nous avons jugé bon de connaître le taux de défaillance de celui-ci.

On dispose ci-dessous des proportions de la variable défaut au sein de notre échantillon :



Valeur	Proportion	%	Comptage
0		82,19	4729
1		17,81	1025

Figure 7:Le taux de défaut dans la base de données.

La proportion des entreprises défaillantes est d'environ 18%. Ce pourcentage est jugé satisfaisant pour la construction d'un modèle basé sur ce critère.

II.4.2.Méthode pas à pas descendante :

La méthode pas à pas descendante se base sur l'introduction de toutes les variables au modèle, puis à éliminer à chaque étape la variable ayant la plus petite contribution dans le modèle.

La dernière étape donne un modèle où toutes les variables sont discriminantes et significatives.

La régression logistique a été effectuée en éliminant la variable « apport*loca » pour ne garder que « apport » vu qu'elles sont corrélées comme il a été cité ci-dessus.

Les résultats de la régression et les tests de validité donnent un meilleur modèle en retenant la variable apport.

Par souci de confidentialité, le tableau obtenu par la régression logistique ne sera pas exposé complètement.

Ci-dessous est un aperçu de ce tableau.

	A	E.S.	Wald	ddl	Sig.	Exp(B)
code_gamme(1)	-,184	,075	6,096	1	,014	,832
code_secteur_act(1)	-,176	,086	4,210	1	,040	,838
code-region			14,059	2	,001	
code-region(1)	-,397	,106	14,058	1	,000	,672
code-region(2)	-,239	,095	6,275	1	,012	,787
categorie_prod_code(1)	-,943	,268	12,425	1	,000	,389
anciennete_Tr			54,522	2	,000	
anciennete_Tr(1)	,674	,091	54,507	1	,000	1,962
anciennete_Tr(2)	,402	,096	17,404	1	,000	1,495
apport%_tr			101,423	2	,000	
apport%_tr(1)	,974	,104	88,319	1	,000	2,648
apport%_tr(2)	,422	,125	11,385	1	,001	1,524
MNT_CAPITAL_Tr(1)	,344	,082	17,738	1	,000	1,411
DUREE_CRED-TR			159,410	2	,000	
DUREE_CRED-TR(1)	,705	,114	38,343	1	,000	2,023

Tableau 29 : Modèle de régression logistique à l'étape 3.

La méthode pas a pas descendante a été effectuée en 3 étapes, et toutes les variables introduites ont été retenues par le modèle sauf les variables code loueur et code société.

Variables hors de l'équation

		Score	ddl	Sig.	
Etape 3(a)	Variables	Code-loueur(1)	2,611	1	,106
		code-soc(1)	,262	1	,609
	Statistiques globales		2,944	2	,230

a. Variable(s) supprimée(s) à l'étape 3 : Code-loueur.

On remarque aussi que toutes les variables retenues ainsi que leurs modalités sont significatives au seuil de 5%.

En effet, les résultats de l'étude statistique ont révélés que ces deux variables ne seraient pas prises par le modèle.

II.5. Contribution des variables retenues :

Le logiciel SPSS Modeler classe les variables du modèle selon l'importance des variables prédites, on obtient le graphe suivant :

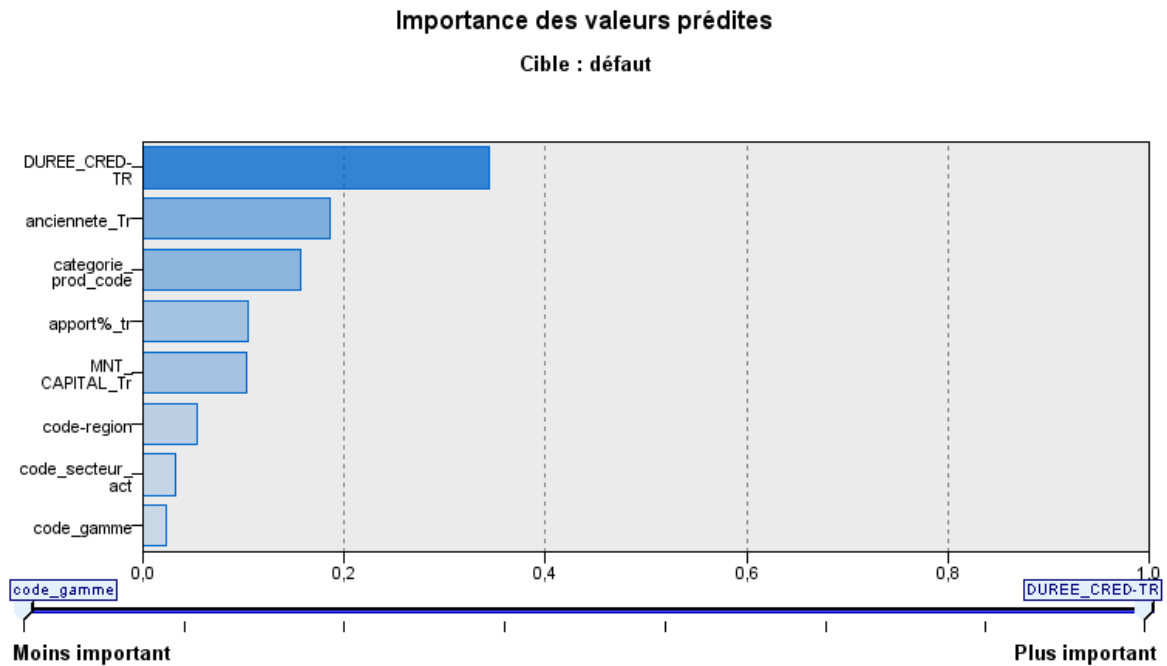


Figure 8: Contributions des variables dans le modèle de la régression logistique

La variable ayant la plus grande contribution est la variable durée de crédit, et celle contribuant le moins dans le modèle est code gamme.

Les contributions en pourcentage associées sont classées comme suit :

<u>Variable</u>	<u>Contribution</u>
durée crédit	33%
Ancienneté	19%
catégorie prod	16%
Apport	11%
Mnt capital	10%
code région	5%
code gamme	3%
code secteur	3%

Tableau 30 : Contributions des variables dans le modèle de la régression logistique

Dans la partie suivante, on va calculer la valeur de l'information dont les résultats devront concorder avec les contributions de chaque variable.

Conclusion

*Après avoir effectué le partitionnement, une étude de corrélation a été effectuée et nous a amené à écarter la variable croisée apport*loca que l'on a trouvé corrélée avec la variable apport.*

Le modèle obtenu par la régression logistique sous SPSS Modeler a finalement retenu 8 variables et a rejeté les 2 variables (Code-loueur et code société). Aussi, toutes les modalités des variables retenues sont significatives au seuil de 5%.

Chapitre 3 : Validation du modèle

Cette partie a pour objectif de valider le modèle de score obtenu par la méthode de régression logistique.

Pour s'assurer de la validité du modèle, plusieurs tests sont efficaces et servent à renseigner sur la qualité et la robustesse du modèle.

I. Cadre théorique.

Il existe plusieurs méthodes pour valider un modèle de scoring, on cite :

I.1. Taux de bon classement :

Le taux de bon classement a pour objectif de valider le pouvoir discriminant des variables de notre modèle via la matrice de confusion qui est formée des entreprises réellement défailtantes et saines, et des entreprises prédites à être défailtantes d'après le modèle conçu.

		Classe prédite		
		Défaillante	Non défaillante	Total
Classe réelle	Défaillante	N1	N2	N1+N2
	Non défaillante	N3	N4	N3+N4
	Total	N1+N3	N2+N4	N1+N2+N3+N4

On distingue 2 taux de bon classement, celui des entreprises saines (TS) et celui des défailtantes (TD) :

$$TD = \frac{N1}{N1 + N2}$$

$$TS = \frac{N4}{N3 + N4}$$

Il est à noter que plus le taux de bon classement s'approche de 1, plus le pouvoir discriminant de nos variables s'avère significatif.

I.2. Courbe de ROC :

La courbe ROC est un outil d'évaluation et de comparaison des modèles. Elle fournit des informations plus intéressantes et plus correctes que celles obtenue par la courbe de Gini ou par le taux de classement dans le sens où elle ne tient pas compte de la proportion de chaque classe dans l'échantillon.

L'indice calculé à partir de la courbe ROC représente l'aire contenue sous la courbe (Area Under the Curve AUC) qui est déduit de l'indice de Gini.

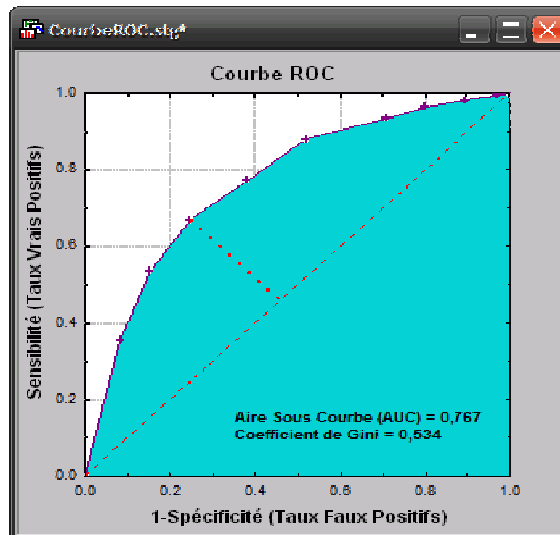


Figure 9: Courbe de ROC

L'aire en bleu est notée A.

Si $\left\{ \begin{array}{l} A=0,5 \quad : \text{ Pas de pouvoir discriminant.} \\ A=1 \quad : \text{ Modèle parfait (fort pouvoir discriminant).} \\ 0,5 < A < 1 : \text{ Pouvoir discriminant raisonnable.} \end{array} \right.$

I.3. Indice de Gini :

La courbe de Gini permet d'illustrer et de calculer le coefficient de Gini (Accuracy Ratio : AR) mesurant les inégalités dans une société donnée. Il représente en général l'inégalité relative au revenu et au patrimoine.

Elle peut facilement être tracée si on dispose de deux échantillons représentatifs de deux classes d'emprunteurs par exemple, dans notre cas il s'agit de clients Défaillants et non-défaillants.

La concavité de cette courbe correspond à une probabilité de défaut conditionnelle issue de la fonction de score construite.

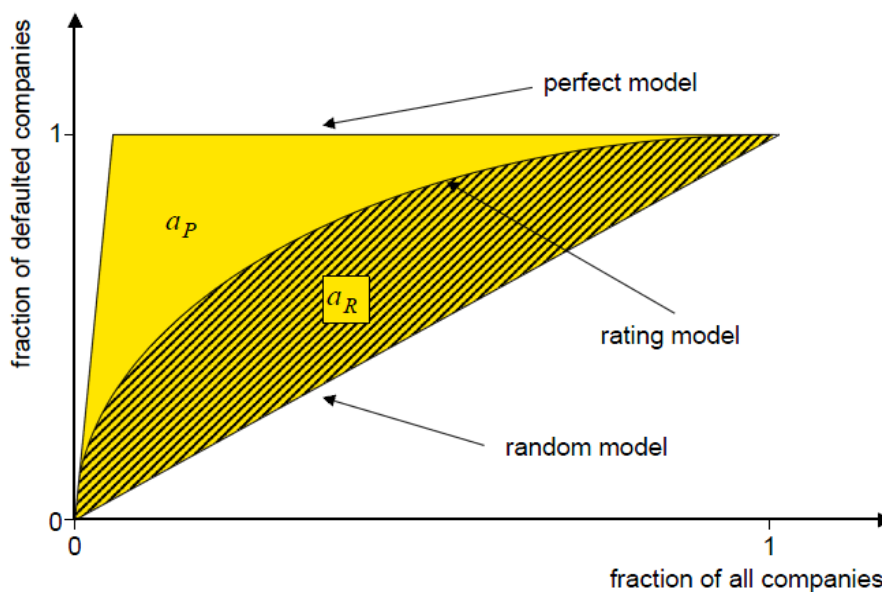


Figure 10: Courbe de Gini

a_R : l'aire entre la courbe CAP du modèle de notation validé et la première bissectrice.

a_P : l'aire entre la courbe CAP du modèle de notation parfait et la première bissectrice.

Ainsi le coefficient de GINI AR est égal:

$$AR = \frac{a_R}{a_P}$$

Plus le coefficient de GINI est proche de 1 et plus le pouvoir discriminant de nos variables est important.

Toutefois il est à noter que la forme de la courbe de Gini dépend de la proportion des bon et mauvais emprunteurs dans l'échantillon.

L'indice de Gini prend généralement des valeurs entre 40% et 70%. Cependant les résultats doivent être interprétés avec prudence du fait qu'ils dépendent énormément de la composition du portefeuille et du nombre de défauts dans l'échantillon.

I.5. Valeur de l'information :

C'est une mesure univariée du caractère prédictif de la variable. Elle mesure la différence de distribution des « bons/mauvais » sur les classes de la variable, pondérée par les effectifs de chaque classe.

La qualité de la variable croît avec la valeur de l'indicateur.

Soit une variable à n modalités. Pour chaque intervalle ou modalité de variable $i = 1, \dots, n$, on définit :

B_i = Nb de Bons de la modalité i .

M_i = Nb de Mauvais de la modalité i .

La valeur de l'information d'une variable est donnée par la formule suivante :

$$v = \sum_{i=1}^n \left[\left(\frac{B_i}{\sum_{i=1}^n B_i} - \frac{M_i}{\sum_{i=1}^n M_i} \right) \times \text{LOG} \left(\frac{B_i \times \sum_{i=1}^n M_i}{M_i \times \sum_{i=1}^n B_i} \right) \right]$$

Remarque : les valeurs sont comprises entre 0 et 3.

II. Cadre applicatif

II.1. Méthode des réseaux de neurones :

II.1.1. Taux de bon classement :

La matrice de confusion nous donne les taux de bon classement obtenus par la méthode des réseaux de neurones.

Les tableaux ci-dessous représentent les taux de classement pour l'échantillon d'apprentissage et l'échantillon test.

✓ Modèle d'apprentissage :

	Prévisible	
	DEFAUT	0.0 1.0
Réal	0.0	80,18%
	1.0	0,83%

Tableau 31: Matrice de confusion en pourcentage pour la variable défaut « échantillon d'apprentissage »

On obtient un taux de bon classement de presque 81,1%. C'est-à-dire que sur 100 clients environs 82 seront correctement classés par le modèle de score.

✓ Modèle test :

	Prévisible	
	DEFAULT	0.0 1.0
Réel	0.0	82,2%
	1.0	1,2%

Tableau 32 : Matrice de confusion en effectif et en pourcentage pour la variable défaut

« échantillon test »

On obtient un taux de bon classement de 83,4%, proche du taux obtenu pour le modèle d'apprentissage.

II.1.2. Indice de GINI :

La deuxième méthode de validation du modèle consiste à calculer la valeur de l'indice de Gini.

Cet indice constitue une bonne mesure de validation de notre modèle de régression logistique.

Dans un premier temps nous avons calculé l'indice de ROC à l'aide du logiciel SPSS.

A partir de la valeur de l'indice de ROC, nous déduisons facilement la valeur de l'indice de Gini.

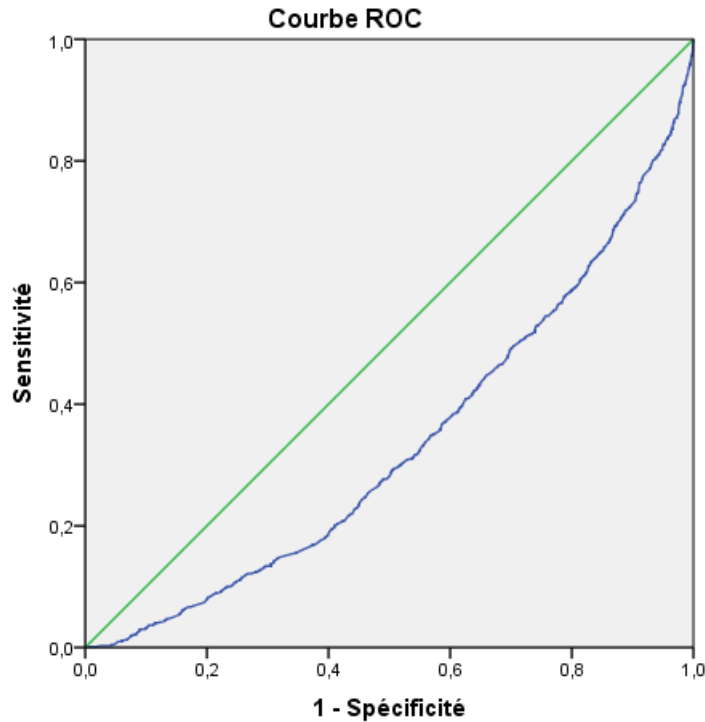
Le calcul de l'indice de ROC donne la valeur de l'indice de GINI à travers la relation suivante.

$$G = 2Z - 1$$

Avec :

- G : indice de GINI.
- Z : indice de ROC

La courbe de ROC du modèle obtenu par la méthode des réseaux de neurones à l'aide SPSS est la suivante :



Les segments diagonaux sont générés par des liaisons.

Et le coefficient de ROC obtenu est :

Zone sous la
courbe

Zone
,336

Ainsi le coefficient de ROC est égale $1 - 0,336 = 66,4\%$ d'où :

Gini = 32,8%

II.2. Méthode de la régression logistique :

II.2.1. Taux de bon classement :

La matrice de confusion donne les taux de bon et de mauvais classement. Les tableaux ci-dessous représentent les effectifs puis les taux de classement pour l'échantillon d'apprentissage et l'échantillon test.

- ✓ Modèle d'apprentissage :

Réal	Prévisible			
	Défaut	0.0	1.0	Total
	0.0	4892	19	4911
1.0	999	38	1037	
Total	5891	57	5948	

Réal	Prévisible		
	DEFAUT	0.0	1.0
	0.0	82,25%	
1.0	0,64%		

Tableau 31: Matrice de confusion en effectif et en pourcentage pour la variable défaut « échantillon d'apprentissage »

On obtient un taux de bon classement de presque 83%. C'est-à-dire que sur 100 clients environs 83 seront correctement classés par le modèle de score.

✓ Modèle test :

Réal	Prévisible			
	DEFAUT	0.0	1.0	Total
	0.0	2148	21	2169
1.0	380	40	420	
Total	2528	61	2589	

Réal	Prévisible		
	DEFAUT	0.0	1.0
	0.0	82,97%	
1.0	1,54%		

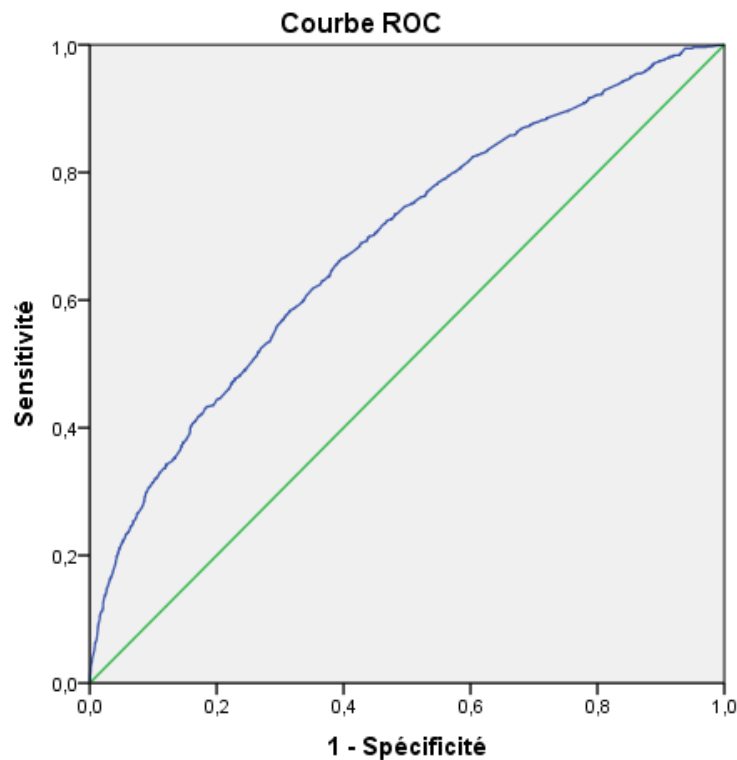
Tableau 32 : Matrice de confusion en effectif et en pourcentage pour la variable défaut

« échantillon test »

On obtient un taux de bon classement de 84,51%, proche du taux obtenu pour le modèle d'apprentissage. Le modèle paraît satisfaisant.

II.2.2. Indice de GINI :

La courbe de ROC du modèle de régression logistique obtenu à l'aide SPSS est la suivante :



Et le coefficient de ROC obtenu est :

Zone sous la courbe

Zone
,687

Ainsi le coefficient de ROC est égale **68,7%**, d'où :

Gini = 37.4%

L'indice de Gini trouvé est de 37.4%. Selon les normes du crédit automobile, dont notre base est essentiellement constituée, il est jugé juste en dessous de la moyenne (moyenne = 40%).

Cet indice s'avère effectivement faible, mais acceptable dans le sens où les données utilisées dans la construction du modèle ne sont pas très pertinentes. En effet, et comme il a été cité dans le chapitre 0, on ne dispose pas des données financières des sociétés comme les ratios financiers, le chiffre d'affaire ou encore les relevés bancaires

II.3. Valeur de l'information :

La valeur de l'information est une mesure du caractère prédictif de la variable.

La formule de calcul est explicitée dans l'annexe IV et donne les résultats suivants :

<u>Variable</u>	<u>Valeur de l'information</u>
durée crédit	0,046
Ancienneté	0,044
catégorie prod	0,039
Apport	0,032
mnt capital	0,026
code région	0,012
code gamme	0,007
code secteur	0,007

On remarque que les variables classées selon la valeur de l'information suivent exactement le même ordre que le classement obtenu par SPSS Modeler et qui calcule la contribution de chaque variable dans le modèle.

<u>Variable</u>	<u>Contribution</u>	<u>Valeur de l'information</u>
durée crédit	33%	0,046
Ancienneté	19%	0,044
catégorie prod	16%	0,039
Apport	11%	0,032
Mnt capital	10%	0,026
code région	5%	0,012
code gamme	3%	0,007
code secteur	3%	0,007

Tableau 33 : Contributions et valeurs de l'information des variables du modèle

Pour ce qui est de la valeur de l'information, après calcul (Voir annexe IV), les valeurs trouvées sont effectivement entre 0 et 3 et il s'avère qu'ils suivent le même ordre de croissance que les contributions des variables.

Notre modèle est donc jugé cohérent.

Conclusion :

La méthode de régression logistique fournit un indice de GINI plus grand. Elle fournit aussi les probabilités de défaillance, chose qu'on ne peut obtenir par le biais de la méthode des réseaux de neurones.

Aussi, la méthode des réseaux de neurones présente certains inconvénients cités dans la partie théorique du chapitre 2 qui concerne la modélisation.

Dans ce qui suit, l'élaboration de la grille de score va se baser sur les résultats obtenus par la régression logistique uniquement.

Chapitre 4 : **Elaboration de la grille de score**

Cette partie a pour objectif de calculer les scores relatifs à chaque modalité des variables qui rentrent dans le modèle. Aussi, on procédera à l'élaboration de la grille de score puis à sa validation.

I . Méthodologie et élaboration de la grille

I.1. Calcul du score :

Le modèle de régression logistique nous a permis d'obtenir les coefficients β de la fonction logistique tel que :

$$\text{Score} = \sum \beta_i X_i.$$

Les X_i représentent les variables explicatives de l'équation dont on a codé les modalités.

On obtient les scores associés à chaque modalité par la formule suivante :

$$\text{Score de la modalité } X_i = \frac{\beta_i - \min(\beta_i)}{\sum (\max(\beta_i) - \min(\beta_i))}$$

Après avoir multiplié chaque score par 1000, on obtient des scores positifs, entre 0 et 1000 et qui croissent avec la diminution du risque.

Ceci veut dire qu'un client très risqué aura un score très bas, et inversement.

Le tableau suivant représente les modalités associées à chaque variable ainsi que le score correspondant :

NB : Suite à une demande de l'organisme d'accueil et par souci de confidentialité, les scores de chaque modalité ont été masqués.

Variables	Modalités	Scores
Code gamme:		
1	Low cost & Haute gamme	XXX
2	Moyenne gamme	XXX
Code secteur:		
1	Commerce détail/ Location/ Services	XXX
2	Industrie/ Transport/ BTP	XXX
Code région:		
1	Meknes/ Casablanca	XXX
2	Fes Saiss / Kenitra Gharb/ Marrakech Haouz/	XXX
	Agadir Sud / Rabat Salé / Autres	
3	Tanger Nord / Eljadida Doukala/ Oujda Orient	XXX

Catégorie produit:			
1	CA		XXX
2	LOA		XXX
Apport:			
1	< de 37%		XXX
2	[37%; 48% [XXX
3	>= 48%		XXX
Montant capital			
1	< 100 000 Dhs		XXX
2	>= 100 000 Dhs		XXX
Durée crédit:			
1	< 36 mois		XXX
2	[36 mois, 60 mois [XXX
3	>= 60 mois		XXX
Ancienneté			
1	< 18 mois		XXX
2	[18 mois; 40 mois [XXX
3	>= 40 mois		XXX

I.2. Grille de score :

A partir de la formule explicitée ci-dessus et à partir de laquelle on a calculé les scores des modalités, on calcule les scores pour chaque observation.

On obtient donc des scores évoluant dans le sens inverse du risque, c'est-à-dire que plus la probabilité de défaillance est élevée, plus le score est bas.

Une fois ces scores obtenus, nous les avons croisés avec la variable de défaut afin de dégager l'effectif et le taux de défaillance pour chaque score. Ci-dessous un aperçu de la procédure suivie :

<i>Comptage</i>	<i>Défaut</i>					
<i>score_global</i>	<i>0</i>	<i>1</i>	<i>Total</i>	<i>%défaut</i>	<i>% prop cumulé</i>	<i>Note</i>
0	3	11	14	78,57%	0,24%	C
38	5	12	17	70,59%	0,52%	C
39	0	4	4	100,00%	0,59%	C
51	7	6	13	46,15%	0,81%	C
58	2	1	3	33,33%	0,86%	C
73	1	2	3	66,67%	0,91%	C
77	1	2	3	66,67%	0,96%	C
85	1	2	3	66,67%	1,01%	C
89	17	15	32	46,88%	1,55%	C
90	2	5	7	71,43%	1,66%	C
96	2	5	7	71,43%	1,78%	C
109	5	1	6	16,67%	1,88%	C
111	9	6	15	40,00%	2,14%	C
112	0	2	2	100,00%	2,17%	C
118	2	2	4	50,00%	2,24%	C
123	7	14	21	66,67%	2,59%	C
124	5	7	12	58,33%	2,79%	C

L'effectif cumulé nous a permis de classer les scores en plusieurs catégories de risque. Ainsi le tableau suivant nous donne les intervalles de score obtenus, la note associée à chaque catégorie de score et le taux de défaut par classe de risque.

<u>Note</u>	<u>Intervalle de score</u>	<u>Effectif</u>	<u>Répartition par intervalle</u>	<u>Effectif défaillance</u>	<u>Taux de défaillance</u>
C	180<	397	4,66%	217	54,66%
B--	[180; 250[414	4,86%	149	35,99%
B-	[250;319[816	9,57%	217	26,59%
B+	[319;494[3501	41,08%	574	16,40%
A-	[494;675[2602	30,53%	253	9,72%
A+	>= 675	793	9,30%	43	5,42%
Total		8523	100%	1453	17,43%

On a dégagé 5 classes de risque, chacune ayant un effectif significatif. Et dont les taux de défaillance décroissent d'une classe à une autre selon la catégorie de risque.

Ainsi, la classe « C » détient le taux de défaut le plus grand, et la classe « A+ », le taux de défaut le plus bas.

Pour effectuer un Scoring, on se restreint aux colonnes suivantes du tableau précédent:

<u>Note</u>	<u>Intervalle de score</u>
C	<180
B--	[180; 250[
B-	[250;319[
B+	[319; 494[
A-	[494 ; 675[
A+	>= 675

Tableau 34 : Système de notation

Ainsi, un client ayant un score < 180 est jugé mauvais et peut se voir refuser une demande de crédit.

II Tests sur la grille de score

La validation temporelle consiste à considérer un nouvel échantillon récent.

Dans notre cas, l'échantillon récent s'étale sur 6 mois (1 er Octobre 2012 → 1 er Avril 2013).

Il s'agit donc d'appliquer les résultats obtenus par le biais de la modélisation théorique sur la base nouvelle pour en déduire des résultats.

Ceci se fera par 2 méthodes qui seront présentées dans ce qui suit :

II.1. Indice de performance (IP) :

II.1.1. Stabilité de la distribution des modalités de chaque variable.

Une des méthodes de validation temporelle de notre modèle consiste à sélectionner ce même nouvel échantillon que l'on a appelé « NEW » et de voir si les proportions de chaque modalité scorée dans cette échantillon ont changées ou au contraire elles sont restées stable par rapport a l'échantillon sur lequel le modèle a été construit et que l'on a nommé « DEV ».

Voici ci-dessous les tableaux et les graphiques qui comparent les proportions de chaque modalité scorée pour chaque variable dans nos deux échantillons :

➤ Code gamme :

Gamme	DEV	NEW
1	39,88%	36,15%
2	60,12%	63,85%

Tableau 35 : proportions des modalités de la variable «code gamme» dans chaque échantillon

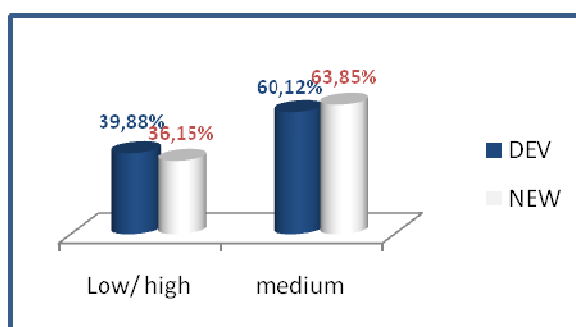


Figure 11: proportions des modalités de la variable «Code gamme» dans chaque échantillon

La figure et le tableau des proportions des modalités de cette variable dans nos deux échantillons, montre que la stabilité des proportions est respectée entre ces deux échantillons.

En effet pour les véhicules Lox cost et Haut de gamme, on est passé d'une proportion de 39,88% dans l'échantillon DEV à 36,15%, et 60,12% à 63,58% pour les véhicules de moyenne gamme.

➤ Code secteur :

secteur	DEV	NEW
1	79,28%	76,71%
2	20,72%	23,29%

Tableau 36 : proportions des modalités de la variable «code secteur » dans chaque échantillon

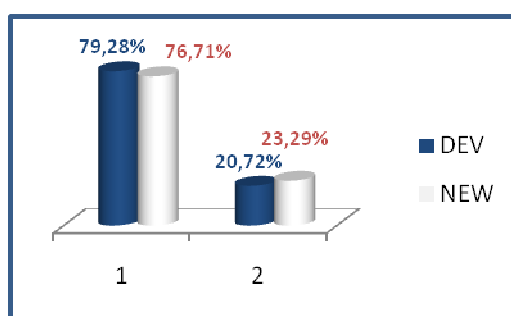


Figure 12: proportions des modalités de la variable «Code secteur » dans chaque échantillon

Pour cette variable on remarque aussi une certaine stabilité temporelle. Pour les secteurs compris dans la classe , on est passé d'une proportion de 79,28% dans l'échantillon DEV à 76,57% dans l'échantillon « NEW », et 20,72% pour les secteurs de la classe 2 à 23,43%.

➤ Catégorie produit :

produit	DEV	NEW
1	3,39%	24,32%
2	96,61%	75,68%

Tableau 37: proportions des modalités de la variable « catégorie produit » dans chaque échantillon

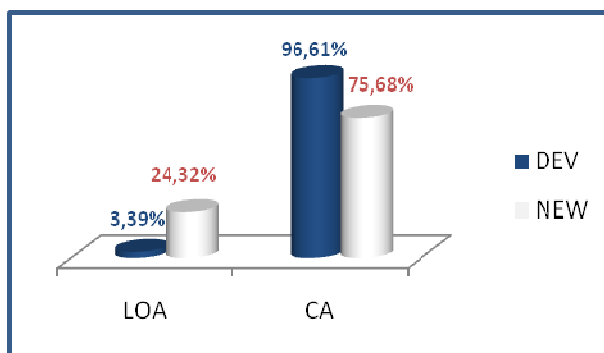


Figure 13: proportions des modalités de la variable « catégorie produit » dans chaque échantillon

La figure et le tableau ci-dessus montrent que pour la variable catégorie de produit, les proportions des deux produits ont significativement changé dans les deux échantillons.

Ceci était prévisible car la modalité 1 qui représente les produits LOA et LOAS concerne des produits nouveaux qui étaient méconnus auprès des clients durant la période de l'échantillon « DEV » mais n'ont été commercialisés que dernièrement et donc commencent à prendre une place dans le marché au profit de la modalité 2 qui représente les crédits Automobile classiques « CA ».

➤ Code région :

région	DEV	NEW
1	32,90%	35,16%
2	50,84%	48,07%
3	16,26%	16,14%

Tableau 38: proportions des modalités de la variable « code région » dans chaque échantillon

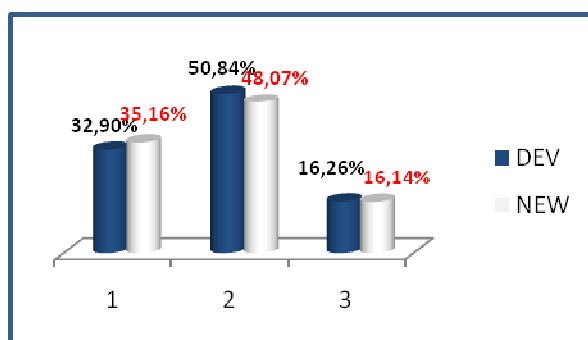


Figure 14: Proportions des modalités de la variable « Code région » dans chaque échantillon

On remarque aussi que pour la variable « code région » les proportions de ses trois modalités n'ont quasiment pas changé entre nos deux échantillons

➤ Ancienneté :

Ancienneté	DEV	NEW
1	34,40%	28,69%
2	25,41%	24,21%
3	40,19%	47,43%

Tableau39 : proportions des modalités de la variable « Ancienneté » dans chaque échantillon

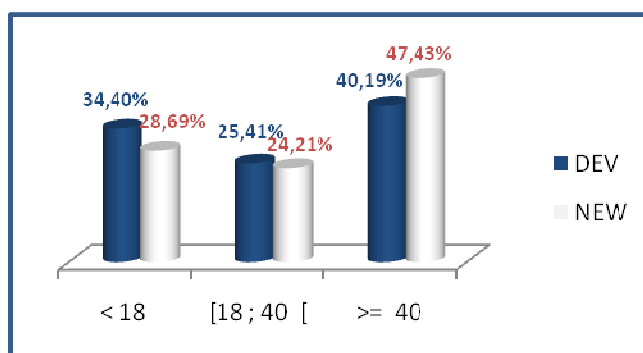


Figure 15: Proportions des modalités de la variable « Ancienneté » dans chaque échantillon.

Pour cette variable on remarque qu'il ya eu un certain changement (Une différence de 7%) entre les proportions des modalités 1 et 3 dans nos deux échantillons.

Ceci s'explique par une maîtrise de risque par le service de prévention risque qui a instauré certaines restrictions relatives à l'ancienneté du client.

➤ Montant capital :

Mnt capital	DEV	NEW
1	25%	31,25%
2	75%	68,75%

Tableau 40 : proportions des modalités de la variable « Mnt capital » dans chaque échantillon

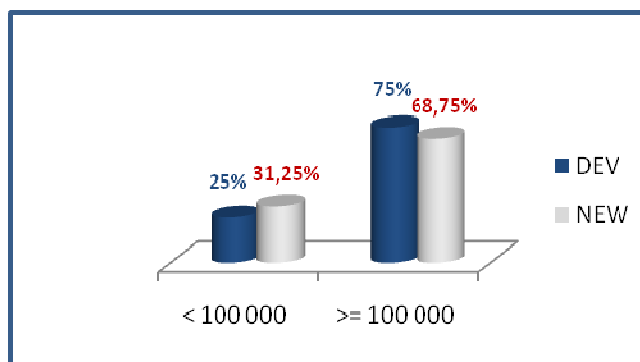


Figure 16: proportions des modalités de la variable « Mnt capital » dans chaque échantillon

La stabilité temporelle des proportions des modalités scorées est confirmée pour la variable montant de capital dont les proportions n'ont que très peu changé.

➤ Apport :

Apport	DEV	NEW
1	58,43%	53,17%
2	20,77%	24,91%
3	20,80%	21,92%

Tableau 41 : proportions des modalités de la variable « Apport » dans chaque échantillon

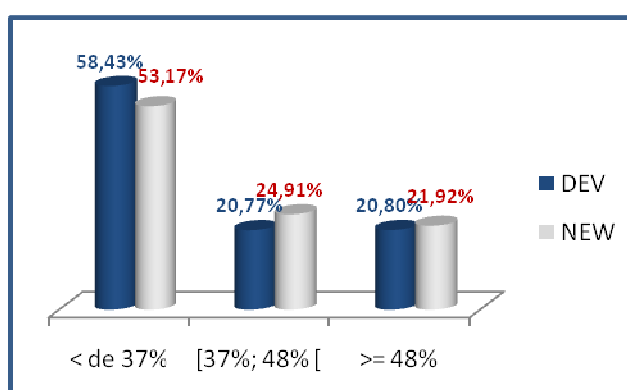


Figure 17: proportions des modalités de la variable « Apport » dans chaque échantillon

On remarque que les proportions des modalités de cette variable ont légèrement changé pour la modalité 1 et 3 (de l'ordre de 5%) tandis que la modalité 2 les proportions sont restées quasi identiques avec une variation de moins de 0,8%.

➤ Durée de crédit :

Durée de crédit	DEV	NEW
1	29,48%	12,97%
2	58,36%	67,17%
3	12,16%	21,86%

Tableau 42 : proportions des modalités de la variable « durée de crédit » dans chaque échantillon

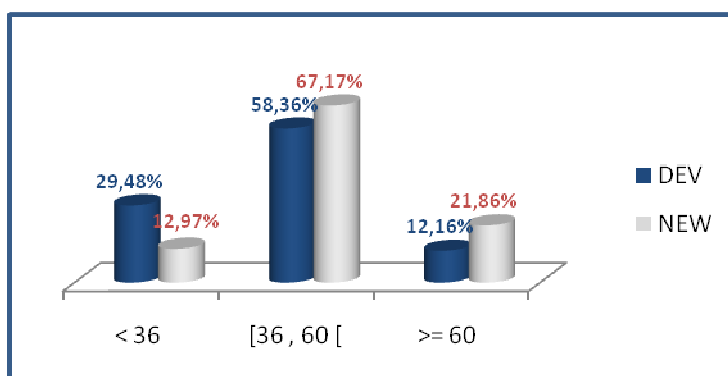


Figure 18: proportions des modalités de la variable « Durée de crédit » dans chaque échantillon

La variable durée de crédit connaît un changement dans la structure de sa population, qui est du à des restrictions imposées par le service en vue de maîtriser le risque.

Le paragraphe suivant va donner la mesure de performance de chaque variable, qui découle du changement de sa structure et de son score.

II.1.2. Indice de performance (IP)

L'indice de performance, comme cité dans la partie théorique concerne chaque variable du modèle. On note :

NEW : échantillon nouveau sur lequel on va valider le modèle.

DEV : échantillon de développement sur lequel le modèle a été construit.

P_{ij} : la proportion de la modalité j de la variable i dans l'échantillon DEV

Q_{ij} : la proportion de la modalité j de la variable i dans l'échantillon NEW

S_{ij} : Score de la modalité j de la variable i .

L'indice de performance (IP) se calcule selon la formule suivante

$$IP_i = \sum_{j=1}^N (Q_{ij} - P_{ij}) S_{ij} * 0.01$$

Tel que N représente le nombre de modalités de la variable i.



Ci-dessous l'indice de performance obtenu pour chacune des variables scorées.

<u>Variable</u>	<u>IP</u>	<u>Stabilité</u>
durée crédit	14,24%	Instable
Ancienneté	4,33%	Légèrement instable
catégorie prod	19,74%	Instable
Apport	3,38%	Légèrement instable
Mnt capital	2,15%	Stable
code région	0,24%	Stable
code gamme	0,69%	Stable
code secteur	0,45%	Stable

On voit que l'indice de performance de la catégorie de produit est de 19.74%, on peut donc conclure que cette variable n'est pas du tout performante. Ceci est dû au fait que « LOA » est un produit nouveau, comme il a été bien cité dans la partie 1 et donc sa commercialisation est récente, ce qui explique les proportions de la variable Catégorie de produit dans le graphique relatif à cette variable.

La variable durée de crédit s'est également avérée instable, ceci est principalement dû aux mesures restrictives imposées par le service prévention risque en vue d'améliorer et de maîtriser le risque.

Ainsi, les durées de crédit supérieures à 36 mois sont privilégiées dans la décision des agents. Ce qui explique le changement dans les proportions.

Pour les variables ancienneté et apport, elles ont également été soumises à certaines restrictions, d'où la légère instabilité.

II.1.3. Test de stabilité de la grille : Kolmogorov-Smirnov

C'est le dernier test à effectuer et il vient après avoir élaboré la grille de notation.

Le test de Kolmogorov-Smirnov a pour but de tester la stabilité du modèle.

Ci-dessous, le tableau du test Kolmogorov-Smirnov obtenu sur :

✓ Défaut=0

Note	Apprentissage			Test			Diff cumulé
	Effectif	%effectif	cumulé	Effectif	%effectif	cumulé	
C	132	2,69%	2,69%	48	2,21%	2,21%	0,47%
B--	188	3,83%	6,52%	87	4,01%	6,22%	0,29%
B-	431	8,78%	15,29%	168	7,75%	13,97%	1,32%
B+	2021	41,15%	56,44%	906	41,77%	55,74%	0,70%
A-	1630	33,19%	89,64%	719	33,15%	88,89%	0,75%
A+	509	10,36%	100,00%	241	11,11%	100,00%	0,00%
Total	4911			2169			

Tableau 43 : Test de Kolmogorov-Smirnov pour les non défaillants

L'écart maximal est $D_{\max} = 1,32$; le seuil de Kolmogorov-Smirnov $D = 3.5$

$$D_{\max} < D$$

Le modèle est jugé stable.

✓ Défaut=1 :

En raison d'effectifs très petits de la population des mauvais, le test ne peut être effectué.

Cependant, on peut nous restreindre à la population des bons clients et à l'ensemble de la population.

✓ Ensemble de la population :

Note	Apprentissage			Test			Diff cumulé
	Effectif	%effectif	cumulé	Effectif	%effectif	Cumulé	
C	286	4,81%	4,81%	111	4,29%	4,29%	0,52%
B--	288	4,84%	9,65%	140	5,41%	9,69%	-0,04%
B-	576	9,68%	19,33%	240	9,27%	18,96%	0,37%
B+	2446	41,12%	60,46%	1055	40,75%	59,71%	0,74%
A-	1815	30,51%	90,97%	787	30,40%	90,11%	0,86%
A+	537	9,03%	100,00%	256	9,89%	100,00%	0,00%
Total	5948			2589			

Tableau 44 : Test de Kolmogorov-Smirnov pour l'ensemble de la population

L'écart maximal est $D_{\max} = 0,86$, qui est inférieur au seuil de Kolmogorov-Smirnov ($D=3.2$).

Ce test nous amène donc à accepter l'hypothèse de stabilité du modèle.

CONCLUSION

Cette partie achève le processus de Score et nous fournit une grille de score adaptée aux très petites entreprises. Une grille qui a été soumise à des tests divers puis validée en conséquence.

La partie suivante consiste en une automatisation du modèle de score obtenu pour pouvoir l'exploiter concrètement



Chapitre 5 :
Automatisation du
modèle de score



I. Automatisation

Dans cette partie nous allons procéder à l'automatisation à l'aide du langage de programmation d'Excel (VBA EXCEL) de la grille de notation précédemment élaborée, afin de faciliter aux agents l'appréciation de la note et du degré de risque de chaque client, et ainsi constituer une aide à la décision quant à l'octroi ou le refus du crédit.

Pour cela il suffit d'ouvrir le fichier Excel d'automatisation et de cliquer dans un premier temps sur le bouton commencer :




Ensuite une interface apparait dans laquelle il faudra saisir les informations relatives à l'entreprise cliente :

Saisie des données

WAFASALAF

Catégorie produit	<input type="text" value="CA"/>	Durée de crédit (Mois)	<input type="text" value="18"/>
Région	<input type="text" value="Fes-Saiss"/>	Apport %	<input type="text" value="40"/>
Secteur d'activité	<input type="text" value="BTP"/> <ul style="list-style-type: none"> Industrie Transport BTP Commerce-Detail Location Services 	Ancienneté (Mois)	<input type="text" value="70"/>
Prix Voiture		Montant de capital (DH)	<input type="text" value="10000000"/>

**Grille de notation pour les TPE:
Crédit Automobile**



وقف سلف
Wafasalaf

Une fois les informations saisies, on clique sur le bouton Calculer et la fenêtre suivante s'affiche :

Note & Appréciation


WAFASALAF

ديما معاك

Note: B+

C	B--	B-	B+	A-	A+
Risque Elevé		Risque Moyen		Risque Faible	

**Grille de notation pour les TPE:
Crédit Automobile**



وقف سلف
Wafasalaf

On obtient donc la note de score.

Pour un client dont la note est dans le rouge ou l'orange, il faudra bien décortiquer le dossier de crédit avant de trancher quant à l'acceptation de la demande.

II. Sauvegarde des informations obtenues.

L'automatisation de la grille de score est un moyen concret d'exploiter les résultats obtenus et donc de constituer une aide à la décision pour les agents.

Cependant, il est important de sauvegarder les résultats relatifs à chaque client.

Ainsi, un bouton sauvegarde a été ajouté au Userform 1, à partir duquel on peut accéder à un fichier Excel qui comporte la liste des modalités choisies ainsi que la note obtenue.

.Pour chaque opération, la liste est incrémentée.

L'Annexe VI représente les résultats obtenus

CONCLUSION GENERALE

Au terme de ce travail, nous avons abouti à un modèle de crédit « scoring » adapté au crédit automobile pour les très petites entreprises (TPE), et qui permet de prédire la catégorie de risque d'un client à travers son score afin d'établir une règle de décision.

Sur la base de données disponible, nous avons appliqué la méthode statistique de la régression logistique. C'est une méthode paramétrique qui consiste à calculer la probabilité de défaillance de chaque client, que l'on transforme en score globale via la fonction logarithmique. Aussi, les paramètres estimés par cette méthode nous ont permis d'établir un score pour chacune des modalités des variables retenues par le modèle.

Nous avons aussi pensé à utiliser d'autres méthodes statistiques comme l'analyse factorielle discriminante (AFD), la méthode des réseaux de neurone pour élaborer la grille de notation. Cette première méthode n'est valable que lorsque toutes nos variables explicatives sont continues, or notre base de données comporte aussi des variables à caractère qualitatif comme la région, le type de société... Quant à la méthode des réseaux de neurones, qui est une méthode non paramétrique, elle présente un indice de GINI inférieur à celui obtenu par la régression logistique, ce qui justifie le choix de l'élaboration de la grille par la régression logistique.

Enfin une application VBA est élaborée en vue d'automatiser la tâche de prévision, et ainsi d'exploiter concrètement la grille de notation élaborée et constituer une aide à la décision pour les agents.

Il est à noter qu'au cours de notre travail, nous avons eu à faire face à quelques contraintes, notamment la non disponibilité de données financières comme les ratios, les mouvements de compte...

Nous espérons que ce travail s'ajoutera à une série d'autres travaux en vue d'établir une base de données plus complète qui améliorera ce modèle.

Finalement, en plus de l'aspect technique, ce projet de fin d'étude nous a permis de découvrir le fonctionnement interne d'un établissement de crédit, d'apprendre la rigueur dans le travail, et de connaître l'esprit d'équipe.

Bibliographie et Webographie

- ✓ Cours d'Analyse des données, Saïd Ramadan 2012/2013
- ✓ Cours d'Analyse d'économétrie avancée, Abdelkhalek Touhami 2012/2013

- ✓ Projet circulaire, approche avancée Bale 2.

- ✓ Data Mining et statistique décisionnelle, Auteur : Stéphane Tufféry.
- ✓ Bernd Engelmann & Robert Rauhmeir. (2006), Springer, The Basel II Risk Parameters (Estimation, Validation, and Stress Testing).

- ✓ Guide IBM SPSS Modeler.
- ✓ Documentation fournie par WAFASALAF
- ✓ Mémoires des Lauréats de l'INSEA traitant le même sujet.

- ✓ www.wikipedia.com
- ✓ www.univ-orleans.fr



ANNEXES



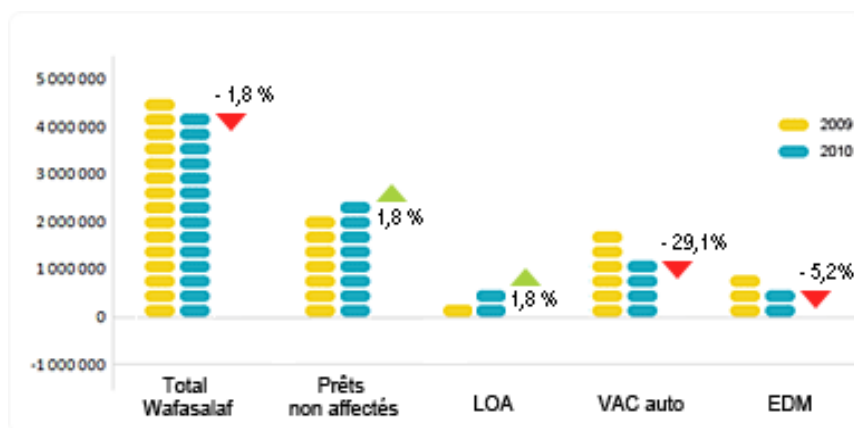
Annexe I : Présentation de l'organisme

Filiale du groupe Attijariwafa Bank, principal acteur bancaire au Maroc.

C'est une Société spécialisée dans le crédit à la consommation, leader sur l'ensemble des compartiments du marché (prêt personnel, crédit automobile, équipement des ménages).

WAFASALAF œuvre par une politique d'innovation et de proximité s'appuyant fortement sur ses domaines d'expertises métiers, à faciliter l'accès au crédit à la consommation et apporter des solutions de financement adaptées aux besoins de ses clients et de ses partenaires. L'ambition de WAFASALAF est de conforter le rang de leader dans l'activité de crédit à la consommation.

Faciliter l'accès au crédit à la consommation, par l'accroissement de notre proximité avec les consommateurs et le renforçant de nos expertises métiers. Le développement de nouveaux canaux de distribution de crédit à la consommation, au moyen de nouvelles technologies de l'information. Le réseau Wafasalaf s'organise autour de :



Evolution Wafasalaf

Acteur spécialisé en crédit à la consommation, Wafasalaf contribue au soutien de l'activité économique en favorisant le développement de la consommation des ménages, par la mise à disposition à la clientèle la plus élargie et de manière responsable des solutions de financement adaptées et innovantes, anticipant l'évolution des habitudes de consommation.

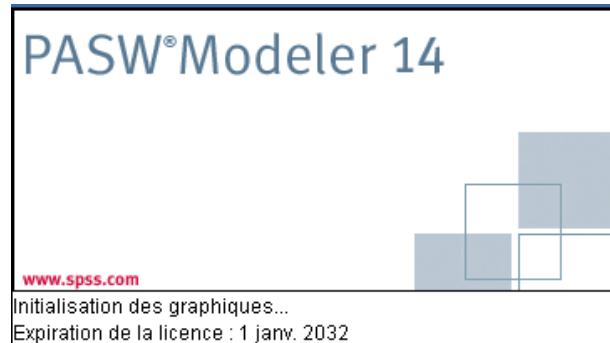
L'activité de traitement du risque se décompose comme suit :

- Prévention et surveillance du risque amont : définition, évaluation de la politique d'acceptation et d'octroi des crédits, sur les différents marchés de Wafasalaf selon ses segments de clientèle,
- Recouvrement amiable et judiciaire : activités de traitement du risque aval, après l'enregistrement d'incidents de paiement par le client,
- Contentieux : cette phase est entamée lorsque les procédures amiables n'aboutissent pas

Wafasalaf a aussi pour activité la gestion pour compte, elle met à disposition de ses partenaires institutionnels (bancaires) son expertise métier, passant de l'étude du risque au recouvrement des clients.

Annexe II : Présentation du logiciel utilisé

✓ SPSS Modeler :

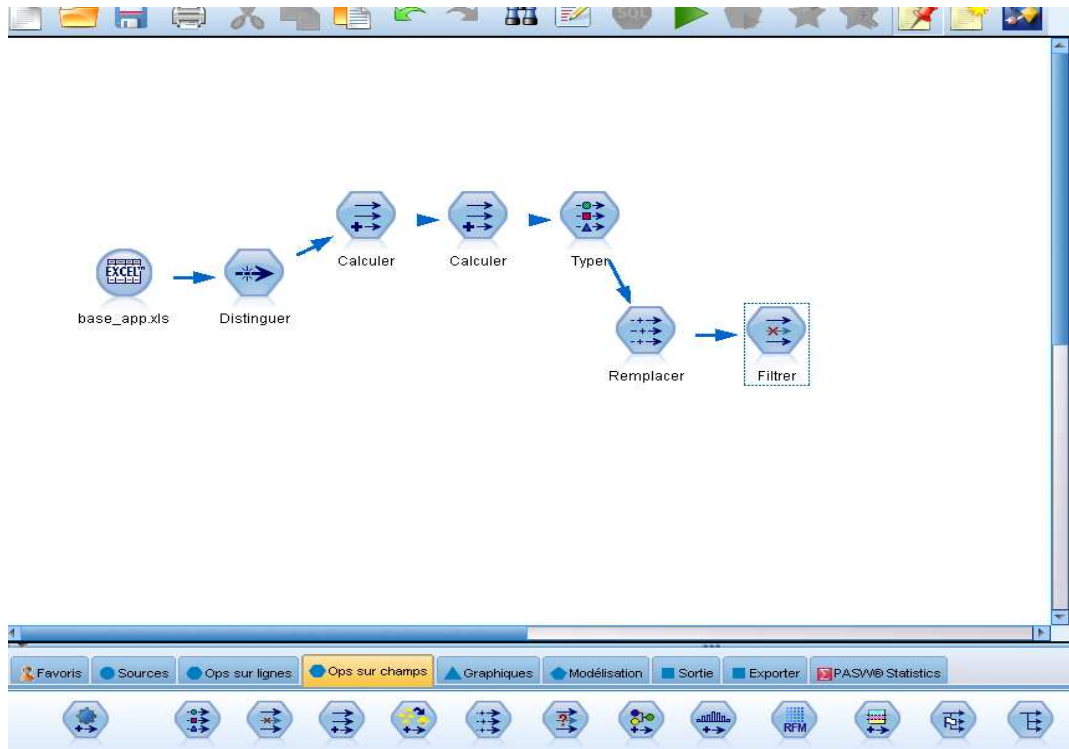


IBM SPSS Modeler est un outil d'exploration des données (data Mining) qui permet de construire des modèles prédictifs sans avoir recours à la programmation et en un temps très court. Il dispose d'une interface visuelle basée sur des fonctions analytiques avancées.

Parmi les fonctionnalités que propose SPSS modeler, nécessaires pour l'élaboration de notre modèle de Scoring, on cite :

- Options de nettoyage pour éliminer ou remplacer les données non valides, imputation automatique des valeurs manquantes et limitation des aberrations et des extrêmes.
- Filtrage des champs, attribution de noms, dérivation, regroupement, recatégorisation, remplacement des valeurs et réorganisation des champs.
- Sélection, échantillonnage (dont échantillonnages classifiés et stratifiés), fusion (dont jointures internes, jointures externes totales, jointures externes partielles et anti jointures) et concaténation des enregistrements ; tri, agrégation et équilibrage.
- Restructuration, partitionnement et transposition des données ...

Ci-dessous, une capture d'écran du flux SPSS Modeler.



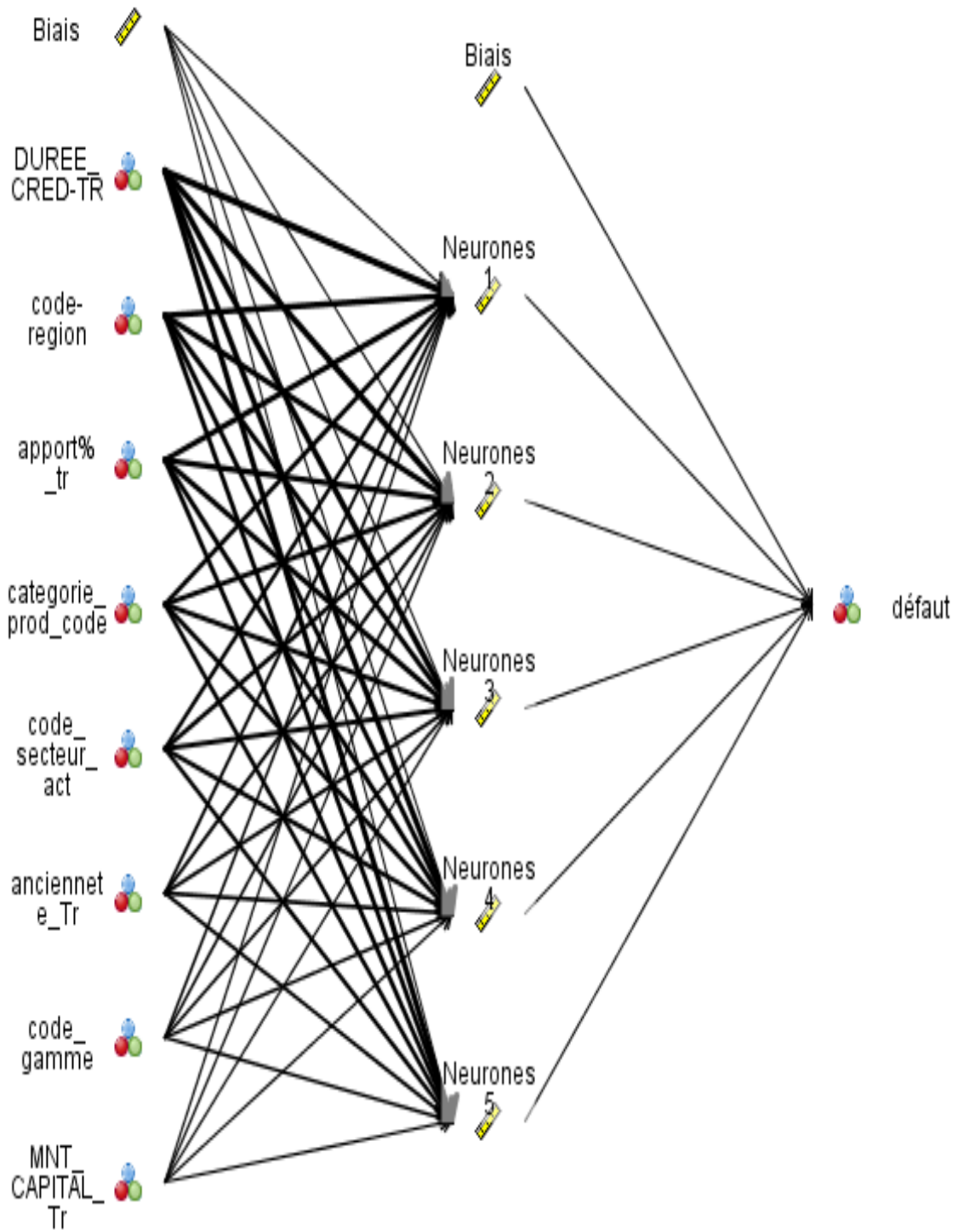
Annexe III : Matrice de corrélation des variables

Correlation Matrix													
	TYPE CO...	SEC ACT...	REGION	MNT CA...	LOUEUR ...	GAMME ...	DUREE ...	CATEGO ...	BANQUE	APPORT ...	APPORT	ANCIENN...	ANCIENN...
TYPE CO...	1.000000	0.016469	-0.001336	-0.005431	0.054472	0.001022	0.032297	0.063158	-0.001904	0.029761	0.015759	-0.005646	-0.013751
SEC ACT...	0.016469	1.000000	-0.151327	0.043395	-0.022878	-0.064053	-0.039248	0.059755	-0.056886	0.002470	0.078076	0.301444	0.044545
REGION	-0.001336	-0.151327	1.000000	-0.046183	0.019089	-0.033670	-0.001988	0.075318	0.019089	0.073425	0.064242	-0.090626	-0.052907
MNT CAPI...	-0.005431	0.043395	-0.046183	1.000000	-0.009002	-0.062189	-0.032265	0.001770	-0.031769	-0.058934	-0.023692	0.157401	0.154134
LOUEUR ...	0.054472	-0.022878	0.019089	-0.009002	1.000000	0.037436	0.085626	0.045440	0.012398	0.049051	0.010690	-0.087625	-0.090682
GAMME P...	0.001022	-0.064053	0.019089	-0.062189	0.037436	1.000000	0.107583	0.026072	0.011449	-0.006799	-0.116140	-0.043434	-0.035290
DUREE C...	0.032297	-0.039248	-0.033670	-0.032265	0.085626	0.107583	1.000000	0.062622	0.008613	-0.116140	0.064100	0.079519	0.079519
CATEGORI...	0.063158	0.059755	-0.001988	0.001770	0.045440	0.026072	0.062622	1.000000	-0.001818	0.081896	0.045314	0.089839	0.083129
BANQUE	-0.001904	-0.056886	0.075318	-0.031769	0.012398	0.011449	0.008613	-0.001818	1.000000	0.029045	0.017225	-0.053233	-0.040646
APPORT L...	0.029761	0.002470	0.073425	-0.058934	0.049051	-0.006799	-0.116140	0.081896	0.029045	1.000000	0.914484	0.914484	-0.172654
APPORT	0.015759	0.078076	0.064242	-0.023692	0.010690	-0.037417	-0.216640	0.045314	0.017225	0.914484	1.000000	-0.158134	-0.199265
ANCIENNE...	-0.005646	0.301444	-0.090626	0.157401	-0.087625	-0.043434	0.064100	0.089839	-0.053233	-0.158134	-0.158975	1.000000	0.945366
ANCIENNETE	-0.013751	0.044545	-0.052907	0.154134	-0.090682	-0.035290	0.079519	0.083129	-0.040646	-0.172654	-0.199265	0.945366	1.000000

Annexe IV : Calcul de la valeur de l'information de quelques variables

Variable	modalités	Bi	Mi	$\sum Bi$	$\sum Mi$	$Bi / \sum Bi$	$Mi / \sum Mi$	$(Bi / \sum Bi) - (Mi / \sum Mi)$	$Bi^2 / \sum Bi$	$Mi^2 / \sum Mi$	$(Bi^2 / \sum Bi) / (Mi^2 / \sum Mi)$	$\text{Log}((Bi^2 / \sum Bi) / (Mi^2 / \sum Mi))$	V
Code gamme	1	1994	369	4911	1037	0,40602729	0,35583414	0,050193149	2067778	1812159	1,141057711	0,05730761	
	2	2917	668			0,59397271	0,64416586	-0,050193149	3024929	3280548	0,922080396	-0,035231211	0,005
Code secteur	1	3954	785	4911	1037	0,80513134	0,75699132	0,048140017	4100298	3855135	1,063593882	0,02677583	
	2	957	252			0,19486866	0,24300868	-0,048140017	992409	1237572	0,801900011	-0,095879781	0,006
Code région	1	1671	293	4911	1037	0,34025657	0,28254581	0,057710762	1732827	1438923	1,204252764	0,080717652	
	2	2483	529			0,50559967	0,51012536	-0,004525687	2574871	2597919	0,991128284	-0,00387013	
	3	757	215			0,15414376	0,20732883	-0,053185074	785009	1055865	0,743474781	-0,128733758	0,012
Catégorie produit	1	354	16	4911	1037	0,07208308	0,01542912	0,056653956	367098	78576	4,671884545	0,66942102	
	2	4557	1021			0,92791692	0,98457088	-0,056653956	4725609	5014131	0,942458225	-0,025737891	0,039
Apport	1	2694	705	4911	1037	0,54856445	0,67984571	-0,131281262	2793678	3462255	0,806895506	-0,093182703	
	2	1179	175			0,2400733	0,16875603	0,071317278	1222623	859425	1,422605812	0,153084579	
	3	1038	157			0,21136225	0,15139826	0,059963984	1076406	771027	1,396067842	0,144906523	0,023

Annexe V : Réseau de neurones PMC (produit)



Annexe VI : Sauvegarde des résultats (VBA Excel)

<u>E Affaire</u>	<u>Catégorie prod</u>	<u>Région</u>	<u>Secteur d'activité</u>	<u>Fin-venture</u>	<u>Durée (ans)</u>	<u>Apport</u>	<u>Financière</u>	<u>Mont capital</u>	<u>Note</u>
DA		El Hadjira-Doukk	Commerce-Detail	30000	57	79	18	850000 A-	
DM		Kenitra-gharb	Transport	13400	50	60	17	1350000 A+	
DA		Kenitra-gharb	Transport	13400	13	40	17	750000 B-	
DA		Oujda	Transport	17800	17	24	17	750000 C	
DA		Oujda	Transport	18500	43	30	45	700000 B+	
DM		Fes-Sais	BTP	18500	43	30	44	700000 A+	
DA		Meknes	BTP	18500	30	57	44	700000 A-	
DM		Fes-Sais	BTP	18500	30	10	44	700000 B+	
DA		Casablanca	Commerce-Detail	15800	20	23	44	700000 B+	
DA		Fes-Sais	Commerce-Detail	15800	34	55	12	70000 B-	
DA		Fes-Sais	Commerce-Detail	15800	34	20	12	70000 C	
DM		Kenitra-gharb	Transport	14000	12	67	45	340000 A+	