

MODÈLE DE FRAUDE

SALAFIN

2019-09-12 v. 03.000

INDICE

1	INTRODUCTION	3
2	MODÉLISATION	4
3	VARIABLES DU MODÈLE	5
3.1	VARIABLES CONTINUES	10
3.1.1	Anc_CB	10
3.1.2	FC	11
3.1.3	Age.Client	12
3.1.4	Montant.Financé	12
3.1.5	DUREE	13
3.1.6	SFDM	13
3.1.7	AGE_GÉRANT	14
3.1.8	SOMME.MENS.CRÉDITS.CONSO.DU.CLIENT	14
3.1.9	MENSUALITÉ	15
3.1.10	REVENU_PARTS	15
3.1.11	CAPITAL.PM.	16
3.1.12	REVENU	16
3.1.13	ANC_EMPLOI	17
3.1.14	PRIX.TTC	17
3.2	Variables catégoriques	18
4	POUVOIR DISCRIMINATION	23
4.1	POUR TYPE CLIENT	23

1 INTRODUCTION

Dans ce document je présente un nouveau modèle prenant en considération les commentaires de L'entreprise par rapport aux premières versions du modèle.

2 MODÉLISATION

La modélisation du modèle a été faite avec l'échantillon élargi proportionné par l'entreprise, qui a un total de 9,966 cas et un pourcentage de fraude d'un 14%. L'échantillon élargi est composé pour les dossiers analysés et identifiés comme fraudes (14%) quelques dossiers analyses et identifiés comme non fraudes, et une sélection aléatoire de dossiers que ne sont pas analysés pour identifier une fraude. Alors, cet échantillon, on peut dire qu'il est représentatif du portefeuille de demandes. Cependant, dans les cas non-analyses pour identifier une fraude pourrie y avoir des dossiers fraudes que nous appelons non-fraudes.

Le modèle a été construit avec une technique de *machine learning* connue sous le nom de *extreme gradient boosting*.

Le modèle obtenu a été cross-validé pour garantir que le modèle ne soit pas trop ajusté aux comportements de l'échantillon de développement.

3 VARIABLES DU MODÈLE

Dans le modèle Il y a un total de 119 variables. La plupart de ces variables représentent une catégorie des variables qui sont catégoriques (le paquet R utilisé a besoin d'une transformation des variables catégoriques en variables *dummies* pour chaque classe).

Une table avec les variables du modèle et quelque commentaires additionnels :

Variable	Commentaires
Anc_CB	Ancienneté bancaire du client au moment de la demande de crédit
FC	Moyenne des flux créditeurs du compte bancaire sur 3 mois
AgeClient	
MontantFinance	
SFDM	Solde de fin de mois du comptes de passif (compte courant, d'épargne, etc) "
ville_BQSIDI.YAHIA.EL.RHARB	
DUREE	
SommeMensCreditsConsoduclient	Somme des mensualités des crédits de consommation en vigueur ou moment de l'octroi"
SecteurActiviteADMINISTRATION.FONCTION.PUBLIQUE	
Age_Gerant	
ville_BQFES	
CodeProduitSD	
Mensualite	Mensualité du crédit demandé
CapitalPM	Capital social des personnes morales
CodeProduitCIM	
AgenceSLFAgence.Casa.Mohamed.V	
REVENU_Parts	Même champ que Revenu à l'exception des professionnels

Variable	Commentaires
	et personnes morales pour lesquels la valeur est nul
ConcessionnaireAutoARIJ.AUTO.sarl	
ville_BQCASABLANCA	
Anc_Emploi	
Revenu	
AgenceSLFDesk.Beni.Mellal	
ConcessionnaireAutoJ.AND.N.OF.AUTOMOBILES	
BanqueCDM	Banque
ville_BQNOUACEUR	
PRIXTTC	Prix du véhicule dans les cas de financement automobile toute taxe comprise
SommeMensCreditsImmoduclient	
ville_BQRABAT	
BanqueBMCE	
ConcessionnaireAutoDELTA.AUTOS.ETOILE	
AgenceSLFAgence.Marrakech	
TypeLogementAutres.logement	
ProfessionArtisan	
TYPE_CLIENTProfessionnels	
AgenceSLFDesk.Côtière	
Ville_TribunalRegistredeCommerceKENITRA	
ProfessionRetraité.Privé	
SecteurActiviteINDUSTRIE	
ville_BQBERRECHID	
ConcessionnaireAutoBOUGHAZ.AUTO	
ville_BQAGADIR	
ProfessionAgriculteur	
AgenceSLFDesk.Oujda	
Anc_ClientAnc	
TypeVehiculeT	
SecteurActiviteBTP	

Variable	Commentaires
MarqueVehiculePEUG	
ConcessionnaireAutoSOPRIAM.PEUGEOT..CASA.	
AgenceSLFDesk.El.Jadida	
ConcessionnaireAutoSUD.CONCESSION.AUTOMOBILE	
ProfessionRetraité	
AgenceSLFAgence.Salé	
AgenceSLFDesk.Kénitra	
BanqueSGMB	
TypeLogementLogé.famille	
Ville_TribunalRegistredeCommerceCASABLANCA	
TypeVehiculeU	
ConcessionnaireAutoWORLD.ISTITMAR	
TYPE_PMLoueur	
MarqueVehiculeCITR	
MarqueVehiculeFIAT	
ville_BQTANGER	
ConcessionnaireAutoSTOKVIS	
Ville_TribunalRegistredeCommerceBERCHID	
ProfessionDirigeant.non.salarie	
MarqueVehiculeFORD	
ConcessionnaireAutoS.G.A.PEUGEOT	
Ville_TribunalRegistredeCommerceTANGER	
SecteurActiviteDIVERS	
SecteurActiviteSERVICES	
TypeLogementPropriétaire	
ProfessionCommerçant	
ville_BQMEKNES	
ville_BQEL.JADIDA	
ville_BQSALE	
MarqueVehiculeVOWA	
MarqueVehiculeSEAT	
Ville_TribunalRegistredeCommerceELJADIDA	
ConcessionnaireAutoTIT.AUTOS	
BanqueCIH	

Variable	Commentaires
TypeVehiculeP	
ProfessionProf.libérale	
ville_BQKENITRA	
ville_BQAIN.HARROUDA	
ville_BQBENI.MELLAL	
BanqueATTIJARIWafa	
ConcessionnaireAutoS.G.A	
SecteurActiviteARTISANAT	
BanqueBP.KENITRA	
ConcessionnaireAutoAUTO.HALL..CASABLANCA.	
AgenceSLFDesk.ALI.YATA	
AgenceSLFDesk.Meknes	
BanqueBP.CASA	
ConcessionnaireAutoJORF.LASFAR.AUTOMOBILE	
BanqueCNCA	
SituationFamilialeM	
SecteurActiviteTourisme...Hôtellerie...Restauration...Loisirs	
ConcessionnaireAutoSCAMA...CASABLANCA	
ville_BQTETOUAN	
MarqueVehiculeRENA	
ProfessionEmployé	
MarqueVehiculeHYUN	
BanqueBP.RABAT	
AgenceSLFDesk.Tétouan	
MarqueVehiculeNISS	
Ville_TribunalRegistredeCommerceAGADIR	
CodeProduitLOA	
ConcessionnaireAutoRENAULT.MAROC	
AgenceSLFAgence.Tanger	
ville_BQTIT.MELLIL	
AgenceSLFAgence.Fes	
ProfessionAgent.de.maitrise...Technicien	
ProfessionAutre	
ConcessionnaireAutoAUTO.HALL..RABAT.	
ConcessionnaireAutoKIA.MOTORS.BINOMIER.GROUP	

Variable	Commentaires
ProfessionCadre.Moyen	
BanqueBP.NORD	
ville_BQMEDIOUNA	
ConcessionnaireAutoETS.HAKAM.FRERES	

Les variables avec leur contribution dans le modèle.

Pour les variables catégoriques, la contribution a été calculé de forme collective avec toutes leurs catégories.

Variable	Gain
Anc_CB	13.39%
ville_BQ	12.06%
FC	9.51%
AgeClient	8.23%
MontantFinance	6.11%
SFDM	5.19%
AgenceSLF	4.78%
ConcessionnaireAuto	4.52%
CodeProduit	4.46%
SecteurActivite	4.39%
DUREE	3.78%
SommeMensCreditsConsoduclient	3.70%
Age_Gerant	3.31%
Mensualite	2.10%
CapitalPM	2.04%
Banque	2.03%
Profession	1.75%
REVENU_Parts	1.66%
Anc_Emploi	1.20%
Revenu	1.17%
MarqueVehiculeCITR	0.88%
TypeLogement	0.84%
PRIXTTC	0.74%
SommeMensCreditsImmoduclient	0.71%

Variable	Gain
TYPE_CLIENT	0.51%
TypeVehicule	0.48%
Anc_Client	0.28%
TYPE_PM	0.15%
SituationFamilialeM	0.04%

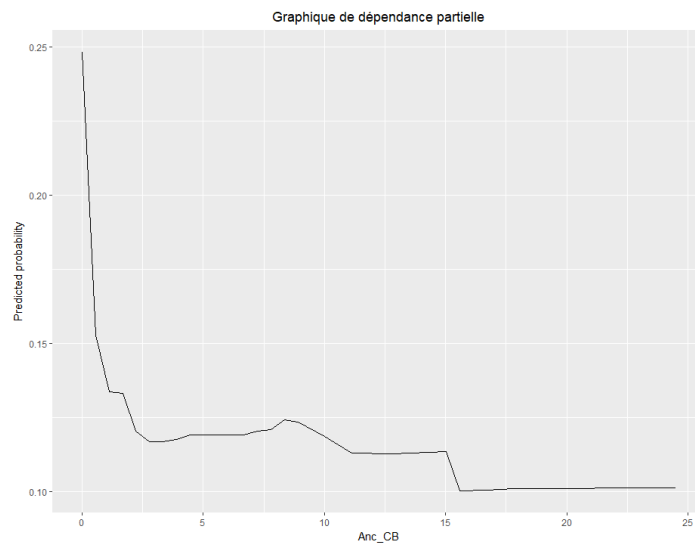
Les variables qui n'ont pas entré dans le modèle :

Variable	Commentaire
Apport.Client....	Cette variable a une tendance à l'envers du sens du business, plus apportation du client, plus risque de fraude. Pour ce motif, cette variable n'est pas dans le modèle.
Nationalité	
TE..Taux.Endettement.	
Sexe	
Autre.Revenu	
Nb_Gerant	

3.1 VARIABLES CONTINUES

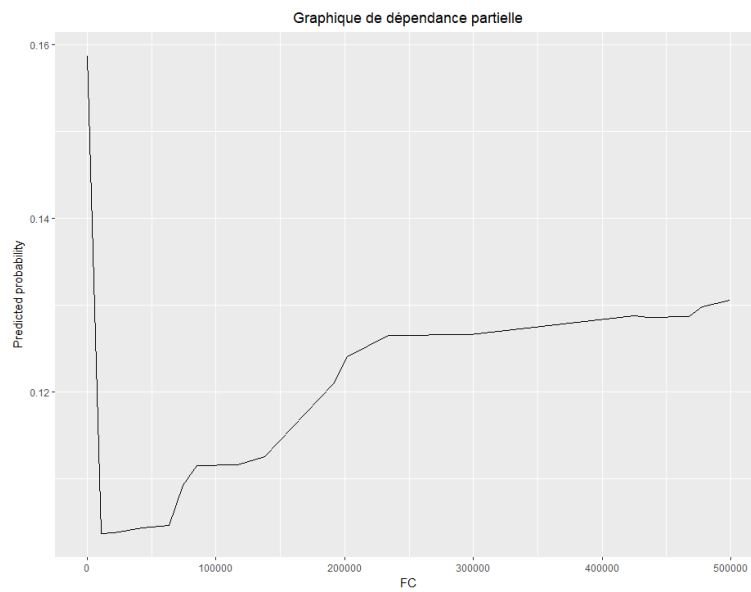
Les tendances des variables continues du modèle.

3.1.1 Anc_CB



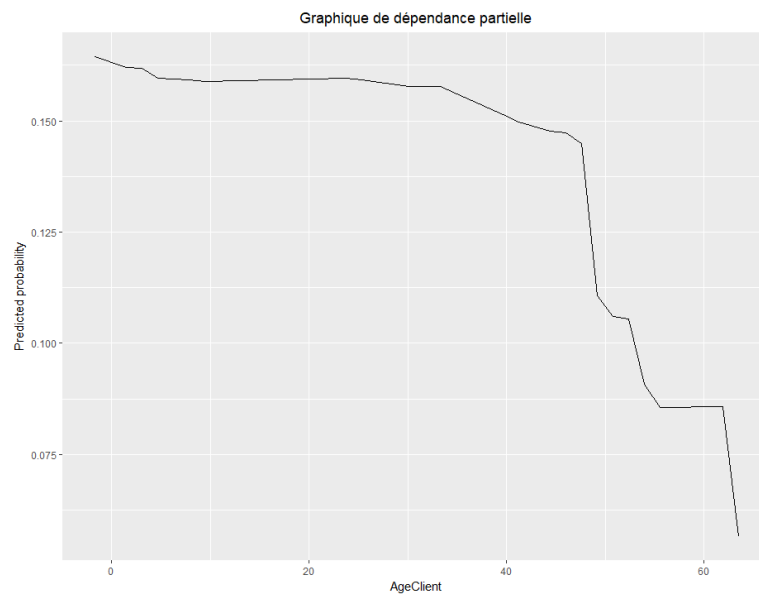
A un valeur plu grande de la variable, moins risque de fraude.

3.1.2 FC



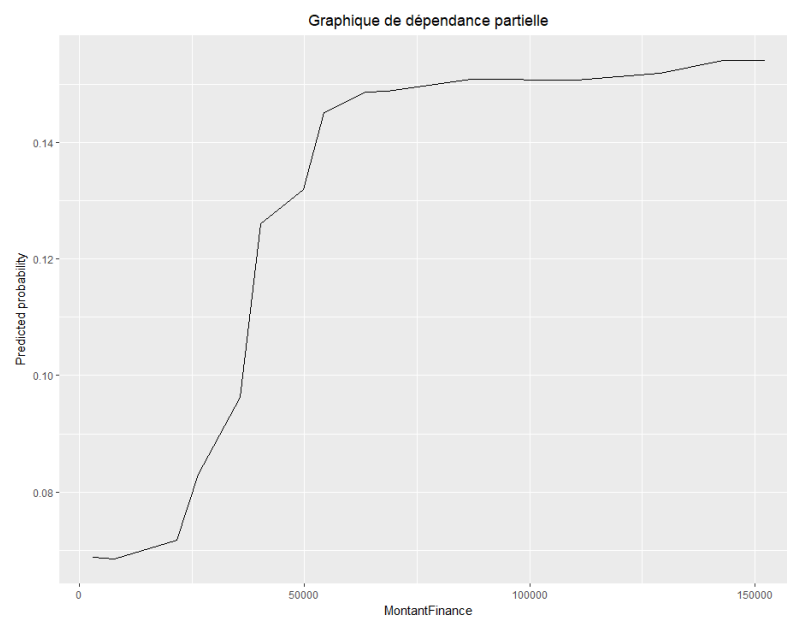
Cette variable a une double tendance.

3.1.3 Age.Client



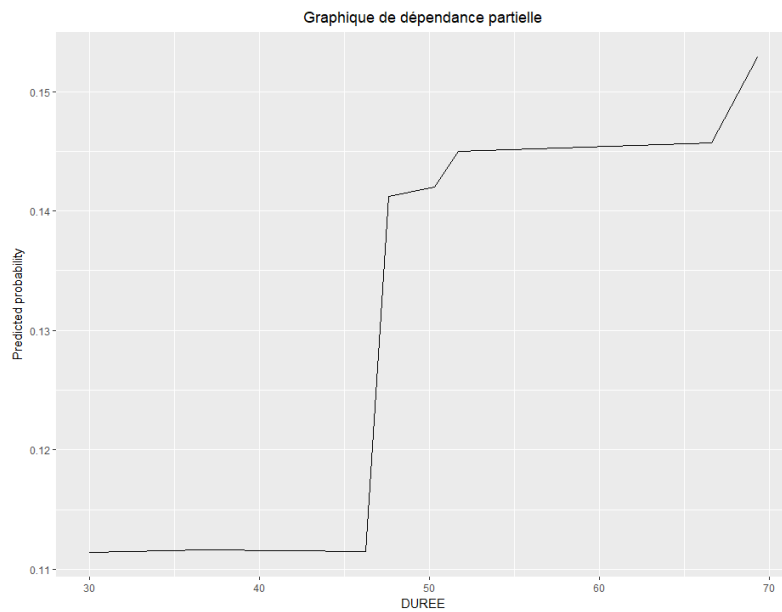
A un valeur plu grande de la variable, moins risque de fraude.

3.1.4 Montant.Financé



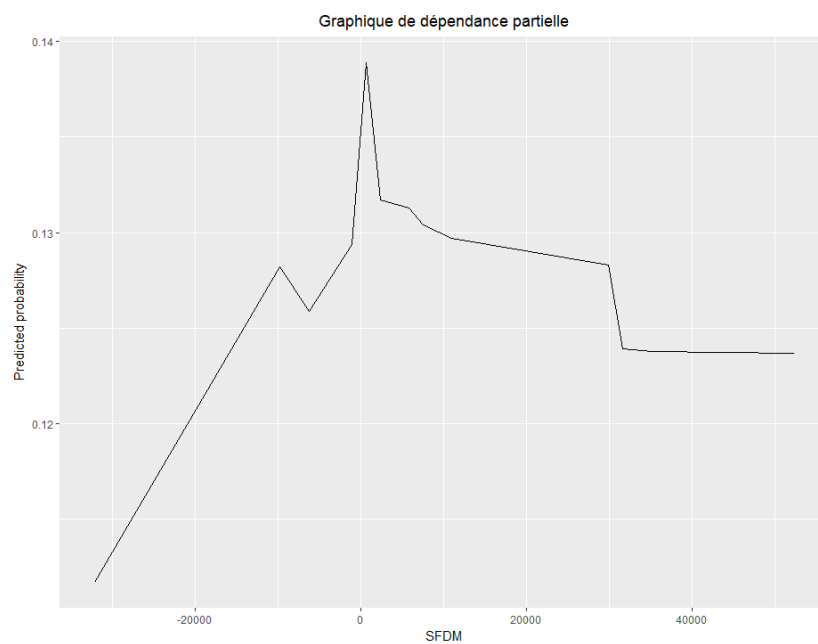
A un valeur plu grande de la variable, plus risque de fraude.

3.1.5 DUREE



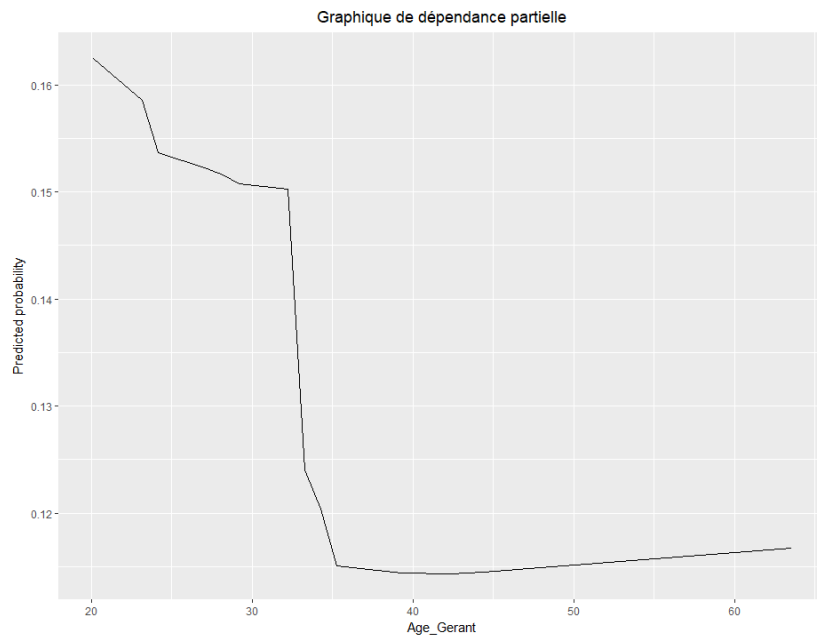
A un valeur plu grande de la variable, plus risque de fraude.

3.1.6 SFDM



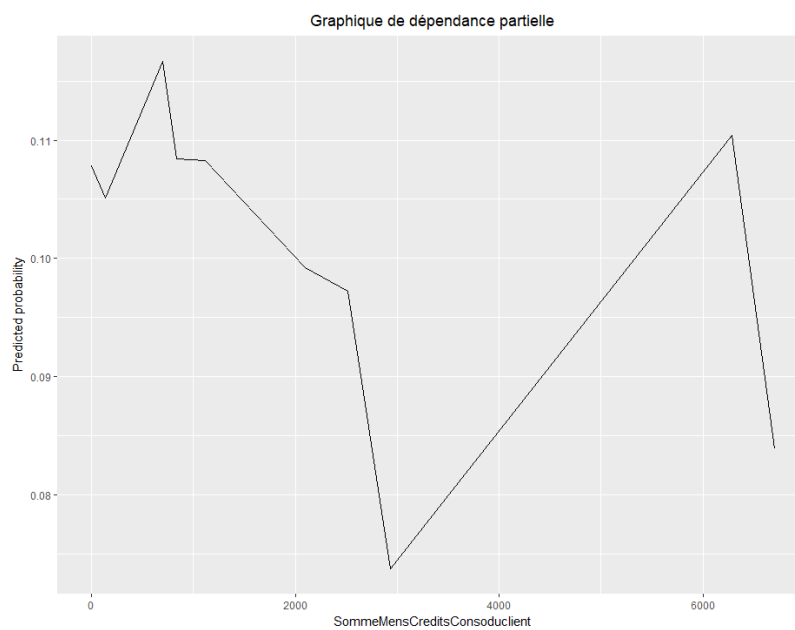
Cette variable a une double tendance.

3.1.7 AGE_GÉRANT



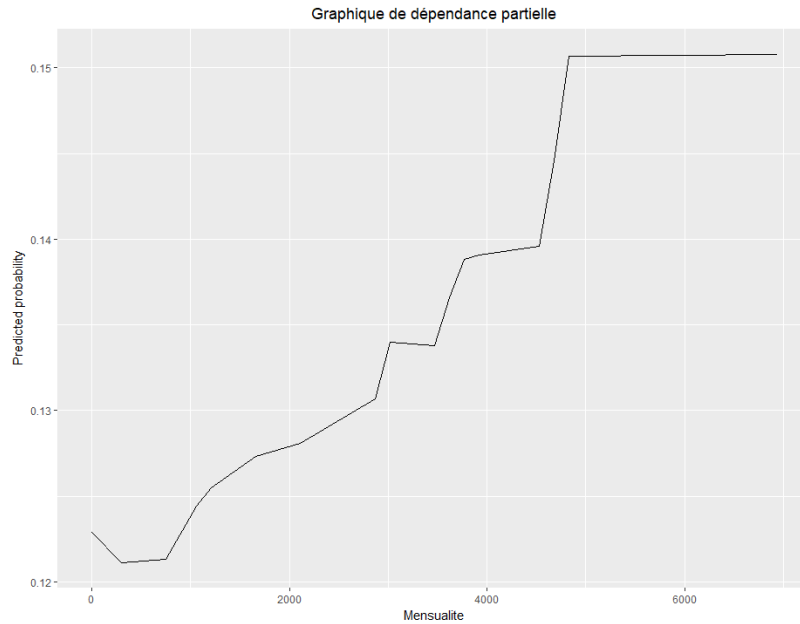
A un valeur plu grande de la variable, moins risque de fraude.

3.1.8 SOMME.MENS.CRÉDITS.CONSO.DU.CLIENT



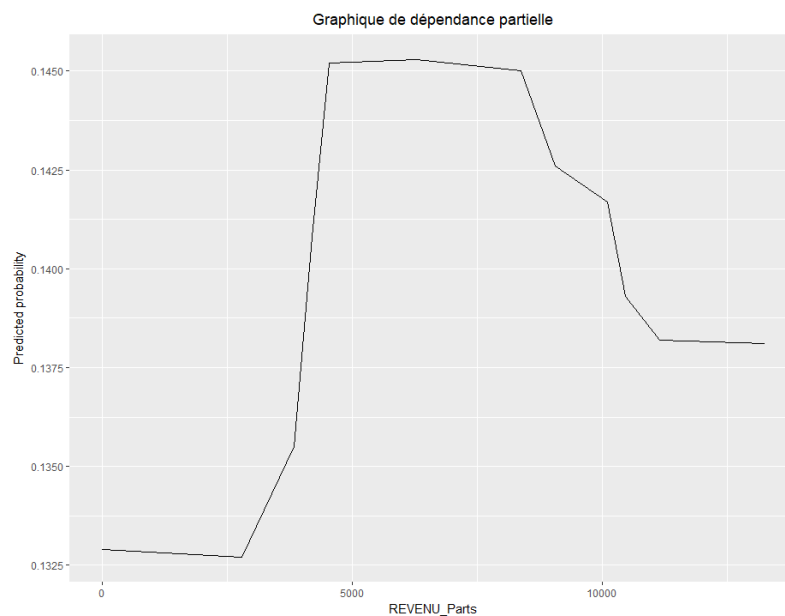
Cette variable a une double tendance.

3.1.9 MENSUALITÉ



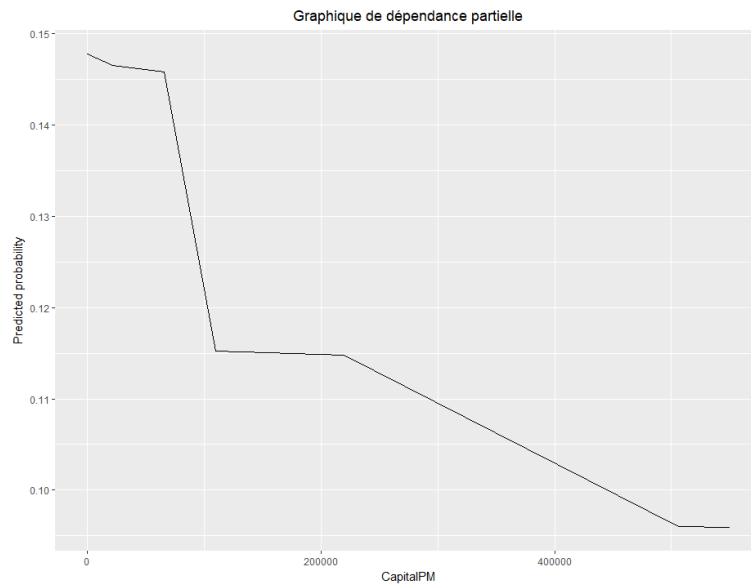
A un valeur plu grande de la variable, plus risque de fraude.

3.1.10 REVENU_PARTS



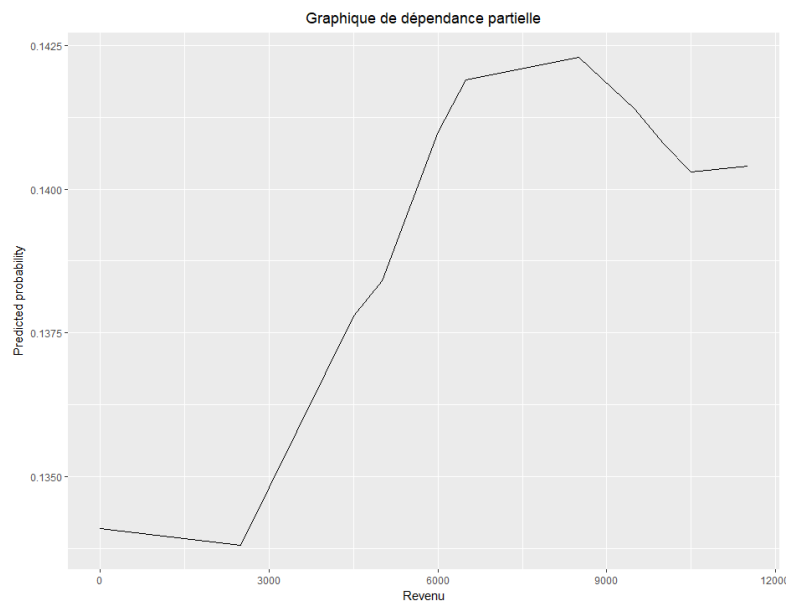
Cette variable a une double tendance.

3.1.11 CAPITAL.PM.



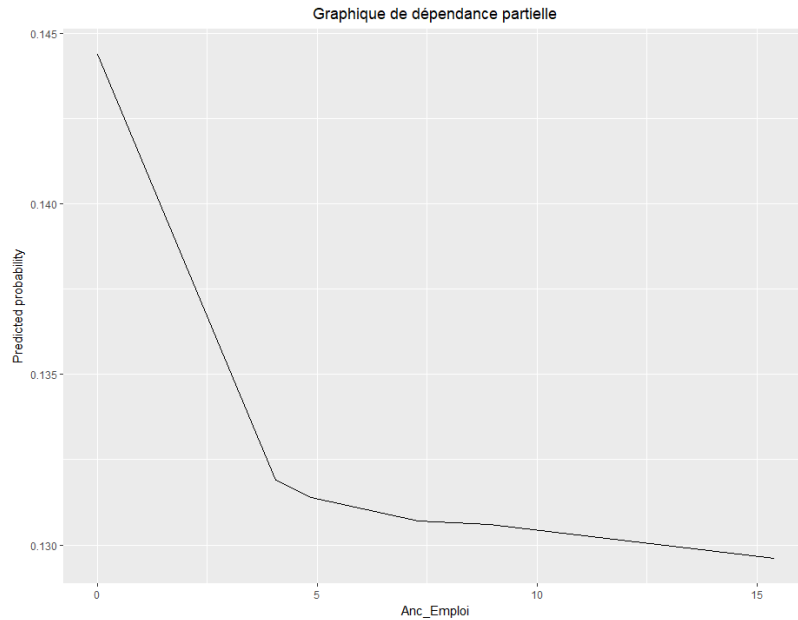
A un valeur plu grande de la variable, moins risque de fraude.

3.1.12 REVENU



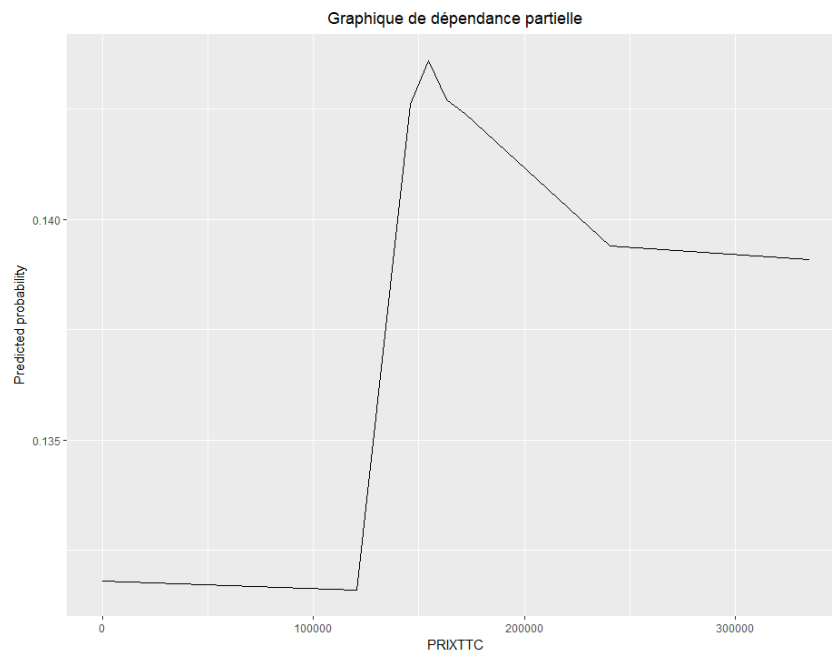
Cette variable a une double tendance.

3.1.13 ANC_EMPLOI



A un valeur plu grande de la variable, moins risque de fraude. La tendance de cette variable a été forcé par AIS.

3.1.14 PRIX.TTC



A un valeur plu grande de la variable, plus risque de fraude.

3.2 VARIABLES CATÉGORIQUES

Les classes de chaque variable catégorique qui participe dans le modèle ordonné selon le risque de fraude (de plus risqué a moins risqué). Il faut que L'entreprise détermine si la tendance de ces variables est correcte.

Ville_TribunalRegistredeCommerce	
Catégorie	Ordonation des catégories (Des plus risqué a moins risqué)
KENITRA	1
CASABLANCA	2
TANGER	3
AGADIR	4
BERCHID	5
ELJADIDA	6

Ville_BQ	
Catégorie	Ordonation des catégories (Des plus risqué a moins risqué)
SIDI.YAHIA.EL.RHARB	1
NOUACEUR	2
FES	3
BERRECHID	4
CASABLANCA	5
AGADIR	6
EL.JADIDA	7
KENITRA	8
TETOUAN	9
AIN.HARROUDA	10
TIT.MELLIL	11
MEDIOUNA	12
TANGER	13
SALE	14

BENI.MELLAL	15
MEKNES	16
RABAT	17

TypeVehicule	
Catégorie	Ordination des catégories (Des plus risqué a moins risqué)
T	1
P	2
U	3

TypeLogement	
Catégorie	Ordination des catégories (Des plus risqué a moins risqué)
Autres.logement	1
Logé.famille	2
Propriétaire	3

SecteurActivite	
Catégorie	Ordination des catégories (Des plus risqué a moins risqué)
DIVERS	1
BTP	2
INDUSTRIE	3
SERVICES	4
Tourisme...Hôtellerie...Restauration...Loisirs	5
ARTISANAT	6
ADMINISTRATION.FONCTION.PUBLIQUE	7

Profession	
Catégorie	Ordination des catégories (Des plus risqué a moins risqué)
Artisan	1
Agent.de.maitrise...Technicien	2
Retraité	3
Commerçant	4
Agriculteur	5
Employé	6
Cadre.Moyen	7
Dirigeant.non.salarié	8
Prof.libérale	9
Autre	10
Retraité.Privé	11

Marque Vehicule	
Catégorie	Ordination des catégories (Des plus risqué a moins risqué)
SEAT	1
FIAT	2
CITR	3
VOWA	4
HYUN	5
RENA	6
PEUG	7
NISS	8
FORD	9

ConcessionnaireAuto	
Catégorie	Ordnation des catégories (Des plus risqué a moins risqué)
J.AND.N.OF.AUTOMOBILES	1
ARIJ.AUTO.sarl	2
TIT.AUTOS	3
S.G.A	4
SOPRIAM.PEUGEOT..CASA.	5
AUTO.HALL..CASABLANCA.	6
SUD.CONCESSION.AUTOMOBILE	7
DELTA.AUTOS.ETOILE	8
JORF.LASFAR.AUTOMOBILE	9
WORLD.ISTITMAR	10
RENAULT.MAROC	11
AUTO.HALL..RABAT.	12
ETS.HAKAM.FRERES	13
KIA.MOTORS.BINOMIER.GROUP	14
STOKVIS	15
S.G.A.PEUGEOT	16
SCAMA...CASABLANCA	17
BOUGHAZ.AUTO	18

CodeProduit	
Catégorie	Ordnation des catégories (Des plus risqué a moins risqué)
SD	1
LOA	2
CIM	3

Banque

Catégorie	Ordination des catégories (Des plus risqué a moins risqué)
CDM	1
CIH	2
BP.CASA	3
BP.RABAT	4
ATTIJARIWafa	5
BP.KENITRA	6
BP.NORD	7
CNCA	8
SGMB	9
BMCE	10

Agences SLF	
Catégorie	Ordination des catégories (Des plus risqué a moins risqué)
Desk.Côtière	1
Agence.Casa.Mohamed.V	2
Desk.ALI.YATA	3
Desk.Kénitra	4
Desk.El.Jadida	5
Desk.Tétouan	6
Agence.Fes	7
Agence.Tanger	8
Desk.Meknes	9
Agence.Salé	10
Desk.Oujda	11
Agence.Marrakech	12
Desk.Beni.Mellal	13

4 POUVOIR DISCRIMINATION

Le pouvoir de discrimination du modèle a été mesuré avec le coefficient de GINI.

Le GINI obtenu dans l'échantillon élargi (d'apprentissage et validation) est d'un 82% et dans l'échantillon réduit d'un 69%. L'échantillon réduit est un échantillon biaisé qui contient les dossiers fraudes et quelque dossier non-fraudes, mais qui sont été analysés manuellement pour écarter la fraude.

4.1 POUR 'TYPE CLIENT'

Le Coefficient du Gini obtenu pour cet model dans les différents types de client est le suivant :

TYPE CLIENT	GINI
Particulier	88.51%
PM	69.51%
Professionnels	75.42%

Si nous faisons un modèle pour chaque segment, nous obtenons les suivants résultats :

TYPE CLIENT	GINI
Particulier	87.66%
PM	67.43%
Professionnels	72.99%

La conclusion est qu'il n'y a pas besoin de faire trois modèles différents, au contraire, c'est un peu meilleur utiliser un modèle global qui a appris de tous les cas disponibles. La méthodologie que nous utilisons est capable de travailler avec différents segments parce que qu'elle est une segmentation lui-même.