Projet « Apprentissage supervisé » Master2 MLSD

Année académique 2019/2020

Enseignant: Lazhar Labiod

Adresse: LIPADE - Université Paris Descartes

Mail: lazhar.labiod@parisdescartes.fr

Objectif

L'objectif de ce travail est la mise en pratique concrète d'un certain nombre de techniques d'apprentissage supervisé (Bayesien Naïf, KNN, LDA, QDA, Linear SVM, Non Linear SVM, Régression logistique, CART et Random Forest), à travers l'étude de données synthétiques et deux cas pratiques nécessitant l'utilisation de logiciels de traitement statistique de données R ou python. Les applications visées sont :

- 1. Carte visa : le scoring d'une base de données de « carte visa ». En résumé, il s'agit de travailler dans ce projet sur une base de données décrivant les clients d'une banque et leurs comportements (mouvements, soldes des différents comptes). L'objectif est l'estimation d'un score d'appétence à la carte VISA Premier. C'est une carte de paiement haut de gamme qui cherche à renforcer le lien de proximité avec la banque en vue de fidéliser une clientèle aisée.
- 2. **Fraude bancaire**: Détection de fraude dans des transactions bancaires: En résumé, il s'agit de travailler dans ce projet sur une base de données décrivant des transactions bancaires sur une période donnée, l'objectif est la détection des transactions frauduleuses.

Je vous encourage à faire preuve d'originalité : vous pouvez très bien utiliser des méthodes qui n'ont pas été présentés en cours, telles que Gradient Boosting, Xgboost et Adaboost.

Etude sur données synthétiques

Cette partie concerne un travail sur données synthétiques. Il s'agit de mettre en œuvre des méthodes vues en cours sur l'ensemble des jeux de données synthétiques proposés. Le travail consiste à réaliser une étude comparative des ces différentes approches de classification supervisée.

Données synthétiques

Il s'agit de 3 bases de données synthétiques possédant des caractéristiques différentes, en termes de nombre de classes et de la structure des classes.

Tables	# d'observations	# de variables	# nombre de classes
Flame	240	2	2
Spiral	312	2	3
Aggregation	788	2	7

Etude de cas pratiques

Cette partie s'intéresse à deux cas pratiques (clients d'une banque, transactions bancaires), l'objectif est d'appliquer les différentes approches vues en cours, choisir pour chaque méthode le meilleur modèle et ensuite comparer ces modèles sur un ensemble de test qui n'a pas été utilisé dans les phases d'apprentissage et de validation des modèles en concurrence.

Données réelles

Data1 : (Visa Premier) : Il s'agit d'une base de données décrivant les clients d'une banque et leurs comportements (mouvements, soldes des différents comptes). La variable à expliquer Y est la variable binaire « Possession de la carte Visa Premier ».

Voici le dictionnaire des variables de la table Visa

matricul departem Département de résidence pivente point de vente sexe Sexe (qualitatif) mbvie sexe Sexe (qualitatif) mitien point de vente sitfamil sitfami	Identif.	Libellé	Identif.	Libellé
privente sexe Sexe (qualitatif) age Age en année sitfamil Situation familiale (Fmar : marié, Fcel : célibataire, Fdiv : divorcé, Fuli : union libre, Fsep : séparé de corps, Fveu : veuf) anciente Anciennet de relation en mois csp Catégorie socio-professionnelle (code num) codeqlt Nombre d'impayés en cours mbimpaye Montant total des rejets en francs mbopguic Nombre d'impayés en cours moycred3 Moyenne des mouvements nets créditeurs des 3 mois en kF aveparmo endette engagemt Total des engagements en francs engagemt mbeptue moysold3 Moyenne des mouvements rediteurs en francs moycred4 Moyenne des souvements curt terme en francs engagemt mbeptue moysold3 Moyenne des mouvements refeiteurs en kF agemut Moyenne des mouvements créditeurs en kF agemut Moyenne des mouvement (en jours) Nombre de produits d'épargne logement mtvie Montant des produits d'épargne logement mteparlt Montant des produits d'épargne logement en francs nbeparlt Montant des produits d'epargne logement en fran	matricul	Matricule (identifiant client)	mtfactur	Montant facturé dans l'année en francs
sexe age Age en année sitfamil Situation familiale (Fmar : marié, Fcel : célibataire, Fdiv : divorcé, Fuli : union libre, Fsep : séparé de corps, Fveu : veuf) anciente Ancienneté de relation en mois Catégorie socio-professionnelle (code num) codeqit Nombre de produits d'épargne monétaire en francs Codeqit Nombre d'impayés en cours. Montant des produits d'épargne logement en francs Montant des comptes sur livret en francs Moyenne des mouvements nets créditeurs des 3 mbeparte mois en kF aveparmo Moyenne des mouvements nets créditeurs des 3 mois en gagemm Total des avoirs épargne monétaire en francs engagemt Total des engagements court terme en francs moyscredi Moyenne des mouvements moyen terme en francs moyeredi Moyenne des mouvements créditeurs en kF aveparmo engagemt Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 motivet motivet Montant des produits d'épargne logement en francs mobivet motivet Montant des produits d'épargne logement en francs nbeparlt Montant des comptes sur livret motivet Montant des produits d'épargne logement en francs nbeparlt Montant des produits d'épargne logement en francs nbeparlt Montant des produits d'épargne logement en francs Nombre de produits d'épargne long terme en francs	departem	Département de résidence	engageml	Engagement long terme
age sitfamil Situation familiale (Fmar : marié, Fcel : célibataire, Fdiv : divorcé, Fuli : union libre, Fsep : séparé de corps. Fveu : veuf) anciente Catégorie socio-professionnelle (code num) codeqlt Code « qualité » client évalué par la banque miteparlo Montant des produits d'épargne logement en francs nombre d'impayés en cours mitivret Montant total des rejets en francs nowberd' moyerne des mouvements nets créditeurs des 3 mois engagemm Total des engagements moyer el agemut moysold3 Moyenne des mouvements moyer de engagemt Moyenne des mouvements créditeurs en francs Nombre de produits épargne monétaire en francs noben notative en francs notative en francs noben notative en francs notative en francs notative en francs noben notative en francs notati	ptvente	Point de vente	nbvie	Nombre de produits contrats vie
Situation familiale (Fmar : marié, Fcel : célibataire, Fdiv : divorcé, Fuli : union libre, Fsep : séparé de corps, Fveu : veuf) Anciennet Ancienneté de relation en mois Catégorie socio-professionnelle (code num) codeqlt Code « qualité » client évalué par la banque mblivret nbimpaye Montant total des rejets en francs Montant total des rejets en francs moycred3 Moyenne des mouvements nets créditeurs des 3 mois moycredi agemut Age du dernier mouvement (en jours) nop Moyenne des mouvements (réditeurs en kF age du dernier mouvement (en jours) nop Montant des produits d'épargne logement en francs ne francs nbeparlt mitvret Montant des comptes sur livret en francs nbeparlt mois en kF aveparmo endette engagemm Total des engagements en francs moycredi agemut Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 Montant des produits d'épargne logement en francs meter de comptes sur livret mointret Montant des produits d'épargne logement en francs nbeparlt Montant des produits d'épargne long terme mitivret meter meter meter meter francs nbeparlt nbeparlt mois en kF aveparmo endette Taux d'endettement mois en kF aveparmo Total des engagements en francs engagemm Total des engagements moyen terme en francs Nombre de produits épargne à terme nbbon Montant des produits bons et certificats en francs Nombre de produits bons et certificats en francs nbeb Nombre de comptes à vue nbecbtar nbecb ave nbecb aver no francs aver per lotal des avoirs sur tous les comptes aver Total des avoirs sur tous les comptes aver Total des avoirs sur tous les comptes aver Total des avoirs épargne financière en francs aver Sexe codé en 0/1 Total des avoirs épargne financière en francs nbecb mouvements créditeurs en kF aver de de mouvements créditeurs en kF aver de de mouvement (en jours) Nombre d'opérations à M-1 Moyenne des mouvements créditeurs en kF aver de de de de menter produits d'épargne long terme nbeparlt Montant des produits d'épargne long terme nbeparlt Montant des produits d'epargne long terme nbeparlt Montant des	sexe	Sexe (qualitatif)	mtvie	Montant des produits contrats vie en francs
taire, Fdiv ; divorcé, Fuli ; union libre, Fsep ; séparé de corps, Fveu ; veuf) anciente Ancienneté de relation en mois Catégorie socio-professionnelle (code num) codeqlt Nombre d'impayés en cours mtivret Montant des produits d'épargne logement en francs nobpatic Nombre d'impayés en cours mtivret Montant total des rejets en francs nobpatic Nombre d'opérations par guichet dans le mois Moyenne des mouvements nets créditeurs des 3 mosenate engagemt Total des engagements en francs engagemt Total des engagements court terme en francs engagemt mbcptvue mysold3 Moyenne des mouvements créditeurs en kF agemut Age du dernier mouvement (en jours) nobparte moverde (code num) nobparte mtivret Montant des comptes sur livret en francs nbeparlt nobparlt Nombre de produits d'épargne long terme en francs nbeparte nbon Nombre de produits d'épargne long terme en francs mteparlt nobparte mover de produits épargne à terme nbon Nombre de produits bons et certificats en francs nbeparte nbon Nombre de produits bons et certificats en francs nbeparte nbon Nombre de produits bons et certificats en francs nbeparte nbon Nombre de paiements par carte bancaire à M-1 Nombre de paiements par carte bancaire à M-1 Nombre de cartes point argent avtscpte aveparfi Total des avoirs épargne financière en francs nbeparte nbon Nombre de avers sur tous les comptes aveparfi Total des avoirs épargne financière en francs nbeparte nouvement (en jours) nombre de cartes voint argent fotal des avoirs épargne financière en francs exer Sexe codé en 0/1 Possession de la carte Visa Premier codé en 0/1	age	Age en année	nbeparmo	Nombre de produits épargne monétaire
paré de corps, Fveu : veuf) Anciente Ancienneté de relation en mois nbeparlo Catégorie socio-professionnelle (code num) codeqlt Code « qualité » client évalué par la banque nblivret Nombre de comptes sur livret Montant des produits d'épargne logement en francs Nombre de comptes sur livret Montant des comptes sur livret nbimpaye mtrejet Montant total des rejets en francs nbeparlt nbopguic Nombre d'opérations par guichet dans le mois moycred. Moyenne des mouvements nets créditeurs des 3 mois en kF aveparmo Total des avoirs épargne monétaire en francs engagemt Total des engagements court terme en francs mbeptive moysold3 Moyenne des soldes moyens sur 3 mois moycredi agemvt Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 Ancienneté de relation en mois mbeparlo Montant des produits d'épargne long terme Montant des produits d'épargne long terme en francs nbeparle Nombre de produits épargne à terme Montant des produits d'épargne long terme en francs nbeparlt Montant des produits d'épargne long terme en francs Nombre de produits d'épargne long terme en francs engagement en francs engagement en francs Nombre de produits d'épargne l	sitfamil	Situation familiale (Fmar : marié, Fcel : céliba-	mteparmo	Montant des produits d'épargne monétaire en
anciente csp Catégorie socio-professionnelle (code num) mteparlo mteparlo mteparlo mteparlo mteparlo momentario momentari		taire, Fdiv : divorcé, Fuli : union libre, Fsep : sé-		francs
codeqlt		paré de corps, Fveu : veuf)		
codeqlt nbimpaye Nombre d'impayés en cours mtlivret Montant des comptes sur livret mtlivret Montant des comptes sur livret montant des produits d'épargne long terme en francs mose en kF aveparmo endete Taux d'endettement mbon Montant des produits épargne à terme mbon Montant des produits bons et certificats mbon Montant des produits bons et certificats en francs mbon Montant des produits bons et certificats en francs nbeptive Nombre de comptes à vue moyald Moyenne des soldes moyen terme en francs mbeptive Montant des produits épargne à terme nbon Montant des produits bons et certificats en francs nbeptive Nombre de produits bons et certificats en francs nbeptive Nombre de paiements par carte bancaire à M-1 nbcb Nombre de paiements par carte bancaire à M-1 nbcb Nombre de cartes point argent avtsepte Total des avoirs sur tous les comptes aveparfi Total des avoirs sur tous les comptes aveparfi Total des avoirs sur tous les comptes aveparfi Total des avoirs épargne financière en francs cartevp Possession de la carte Visa Premier codé en 0/1 cartevpr	anciente	Ancienneté de relation en mois	nbeparlo	Nombre de produits d'épargne logement
nbimpaye mtrejet Montant total des rejets en francs nbopguic Nombre d'opérations par guichet dans le mois moycred3 Moyenne des mouvements nets créditeurs des 3 nbeparte mois en kF aveparmo endette engagemt Total des engagements en francs engagemt Total des engagements court terme en francs engagemm nbeptvue moysold3 Moyenne des mouvements nets créditeurs en francs nbopaic Nombre d'opérations par guichet dans le mois miteparte mois en kF Total des avoirs épargne monétaire en francs engagemt Total des engagements court terme en francs engagemm nbeptvue moysold3 Moyenne des soldes moyens sur 3 mois moyeredi Moyenne des mouvements créditeurs en kF age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 Total des rejets en francs mteparlt moteparlt moteparle moteparlt moteparle moteparl	csp	Catégorie socio-professionnelle (code num)	mteparlo	그렇는 것 맛있는 하는데 10대에 가장 하시다는 사람들은 사람들이 가장하게 되었다.
mtrejet Nombre d'opérations par guichet dans le mois mteparlt nbopguic Nombre d'opérations par guichet dans le mois mteparlt Montant des produits d'épargne long terme en francs moyered Moyenne des mouvements nets créditeurs des 3 mois en kF aveparmo aveparmo rotal des avoirs épargne monétaire en francs endette averagement Total des engagements en francs engagemt Total des engagements court terme en francs engagemm nbeptvue Nombre de comptes à vue moysold Moyenne des soldes moyens sur 3 mois moyeredi Moyenne des mouvements créditeurs en kF agemvt Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 moyente dans le mois mteparlt metal des produits d'épargne long terme Montant des produits d'épargne long terme en francs nbeparte Nombre de produits épargne à terme nbebon Nombre de produits épargne à terme nbebon Nombre de produits bons et certificats en francs Montant des produits bons et certificats en francs nbebon Nombre de paiements par carte bancaire à M-1 nbeb Nombre de cartes point argent avtsepte aveparfi Total des avoirs sur tous les comptes aveparfi Total des avoirs épargne financière en francs Possession de la carte Visa Premier codé en 0/1 Possession de la carte Visa Premier codé en 0/1	codeqlt	Code « qualité » client évalué par la banque	nblivret	Nombre de comptes sur livret
moycred3 Moyenne des mouvements nets créditeurs des 3 mois en kF aveparmo endette Taux d'endettement engagemt Total des engagements court terme en francs engagemt mobre de comptes à vue moysold3 Moyenne des soldes moyens sur 3 mois moycredi Moyenne des mouvement (en jours) nbop Nombre d'opérations à M-1 moycred3 Moyenne des mouvement des créditeurs des 3 mois moyene de comptes à vue moyened in Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 moycred3 Moyenne des soldes moyens sur 3 mois moyene des mouvement (en jours) nbop Nombre d'opérations à M-1 moycred3 Moyenne des soldes moyens sur 3 mois moyene des mouvement (en jours) nbop Nombre d'opérations à M-1 moycred3 moyene des soldes moyens sur 3 mois moyene d'opérations à M-1 moycred4 moyene d'opérations à M-1 moyene d'opérations à M-1 moyene d'opérations à M-1 moyene des mouvement (en jours) moyene de paiements par carte bancaire à M-1 nbcb Nombre de cartes point argent avtscret aveparfi Total des avoirs sur tous les comptes aveparfi Total des avoirs épargne financière en francs Sexe codé en 0/1 Possession de la carte Visa Premier codé en 0/1	nbimpaye	Nombre d'impayés en cours	mtlivret	Montant des comptes sur livret en francs
moycred3 Moyenne des mouvements nets créditeurs des 3 nbeparte mois en kF aveparmo endette Taux d'endettement engagemt Total des engagements en francs engagemc engagemc nouvements nets créditeurs en francs engagemm nbeptvue Nombre de comptes à vue moysold3 Moyenne des soldes moyens sur 3 mois moyeredi agemvt Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 Montant des produits épargne à terme Montant des produits bons et certificats en francs mtbon Montant des produits bons et certificats en francs nbon Montant des produits bons et certificats en francs nbon Montant des produits bons et certificats en francs nbon Montant des produits bons et certificats en francs nbon Montant des produits bons et certificats en francs nbon Montant des produits bons et certificats en francs nbon Montant des produits bons et certificats en francs nbon Montant des produits bons et certificats en francs nbon Montant des produits bons et certificats en francs nbon Montant des produits bons et certificats en francs nbon Montant des produits épargne à terme Montant des produits épargne à terme Nombre de produits épargne à terme nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de produits bons et certificats en francs nbon Nombre de paiements pa	mtrejet	Montant total des rejets en francs	nbeparlt	Nombre de produits d'épargne long terme
mois en kF aveparmo endette Taux d'endettement engagemt Total des engagements en francs engagemc engagemc Total des engagements court terme en francs engagemm nbcptvue Nombre de comptes à vue moysold3 Moyenne des soldes moyens sur 3 mois moycredi Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 mois en kF Total des avoirs épargne monétaire en francs en francs en francs nbbon Montant des produits bons et certificats en francs nbpaiecb Nombre de paiements par carte bancaire à M-1 Nombre de cartes point argent avtscpte Total des avoirs sur tous les comptes aveparfi Total des avoirs épargne financière en francs agemvt Age du dernier mouvement (en jours) Nombre d'opérations à M-1 Total des engagements moyen terme en francs nbcb Nombre de cartes point argent avtscpte Total des avoirs sur tous les comptes aveparfi Total des avoirs épargne financière en francs cartevp Possession de la carte Visa Premier codé en 0/1 Possession de la carte Visa Premier codé en 0/1	nbopguic	Nombre d'opérations par guichet dans le mois	mteparlt	
engagemt Total des engagements en francs engagemm Total des engagements court terme en francs engagemm nbcptvue Nombre de comptes à vue moysold3 Moyenne des soldes moyens sur 3 mois moycredi Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 Total des engagements court terme en francs nbcptvue Nombre de paiements par carte bancaire à M-1 nbcb Nombre de paiements par carte bancaire à M-1 nbcb Nombre de cartes point argent avtscpte aveparfi Total des avoirs sur tous les comptes aveparfi Total des avoirs épargne financière en francs cartevp Possession de la carte Visa Premier Sexe codé en 0/1 Possession de la carte Visa Premier codé en 0/1	moyered3		nbeparte	Nombre de produits épargne à terme
engagemt Total des engagements en francs engagemm Total des engagements court terme en francs engagemm nbcptvue Nombre de comptes à vue moysold3 Moyenne des soldes moyens sur 3 mois moycredi Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 Total des engagements court terme en francs nbcptvue Nombre de paiements par carte bancaire à M-1 nbcb Nombre de paiements par carte bancaire à M-1 nbcb Nombre de cartes point argent avtscpte aveparfi Total des avoirs sur tous les comptes aveparfi Total des avoirs épargne financière en francs cartevp Possession de la carte Visa Premier Sexe codé en 0/1 Possession de la carte Visa Premier codé en 0/1	aveparmo	Total des avoirs épargne monétaire en francs	mteparte	Montant des produits épargne à terme
engagemm Total des engagements court terme en francs engagemm nbeptivue Mombre de comptes à vue moysold3 Moyenne des soldes moyens sur 3 mois moyeredi Age du dernier mouvement (en jours) nbop Nombre d'opérations à M-1 Francs Nombre de paiements par carte bancaire à M-1 Nombre de cartes point argent avtsepte aveparfii Total des avoirs sur tous les comptes aveparfii Total des avoirs épargne financière en francs cartevp Possession de la carte Visa Premier sexer Sexe codé en 0/1 Possession de la carte Visa Premier codé en 0/1 Possession de la carte Visa Premier codé en 0/1	endette		nbbon	
engagemm nbeptvue Nombre de comptes à vue moysold3 Moyenne des soldes moyens sur 3 mois moyeredi agemvt nbop Nombre d'opérations à M-1 Total des engagements moyen terme en francs nbeb Nombre total de cartes Nombre de cartes point argent Total des avoirs sur tous les comptes aveparfi Total des avoirs sur tous les comptes aveparfi Total des avoirs épargne financière en francs cartevp Possession de la carte Visa Premier Sexe codé en 0/1 Possession de la carte Visa Premier codé en 0/1	engagemt	Total des engagements en francs	mtbon	- Table
nbeptvue Nombre de comptes à vue nbeptar moysold3 Moyenne des soldes moyens sur 3 mois moyeredi Moyenne des mouvements créditeurs en kF agemvt Age du dernier mouvement (en jours) Nombre d'opérations à M-1 sexer Sexe codé en 0/1 Possession de la carte Visa Premier codé en 0/1	engageme	Total des engagements court terme en francs	nbpaiecb	Nombre de paiements par carte bancaire à M-1
moysold3 Moyenne des soldes moyens sur 3 mois moycredi Moyenne des mouvements créditeurs en kF agemvt Age du dernier mouvement (en jours) Nombre d'opérations à M-1 Sexer Sexer Cartevpr Possession de la carte Visa Premier codé en 0/1 Possession de la carte Visa Premier codé en 0/1	engagemm	Total des engagements moyen terme en francs	nbeb	Nombre total de cartes
moysold3 Moyenne des soldes moyens sur 3 mois moycredi Moyenne des mouvements créditeurs en kF agemvt Age du dernier mouvement (en jours) Nombre d'opérations à M-1 Sexer Sexer Cartevpr Possession de la carte Visa Premier codé en 0/1 Possession de la carte Visa Premier codé en 0/1	nbcptvue	Nombre de comptes à vue	nbebptar	Nombre de cartes point argent
agemyt Age du dernier mouvement (en jours) cartevp nbop Nombre d'opérations à M-1 sexer Sexe codé en 0/1 cartevpr Possession de la carte Visa Premier Possession de la carte Visa Premier codé en 0/1	moysold3	Moyenne des soldes moyens sur 3 mois	avtsepte	Total des avoirs sur tous les comptes
nbop Nombre d'opérations à M-1 sexer Sexe codé en 0/1 cartevpr Possession de la carte Visa Premier codé en 0/1	moycredi	Moyenne des mouvements créditeurs en kF	aveparfi	Total des avoirs épargne financière en francs
cartevpr Possession de la carte Visa Premier codé en 0/1	agemyt	Age du dernier mouvement (en jours)	cartevp	Possession de la carte Visa Premier
0/1	nbop	Nombre d'opérations à M-1	sexer	Sexe codé en 0/1
nbjdebit Nombre de jours de débit			cartevpr	
			nbjdebit	Nombre de jours de débit

Data2: Credit card_Fraud: (pour plus de détails, voir https://www.kaggle.com/dalpozz/creditcardfraud).

Le jeu de données contient les transactions effectuées par cartes de crédit en septembre 2013 par les titulaires de carte européens. Cet ensemble de données présente les transactions qui se sont produites en deux jours, où nous avons 492 fraudes sur 284 807 transactions. L'ensemble de données est très déséquilibré, les classes positives (fraudes) représentent 0,172% de toutes les transactions. Il contient uniquement des variables d'entrée numériques résultant d'une transformation PCA. Les caractéristiques V1, V2, ... V28 sont les composantes principales obtenues avec PCA, les seules caractéristiques qui n'ont pas été transformées avec PCA sont 'Time' et 'Amount'. La variable 'Time' contient les secondes écoulées entre chaque transaction et la première transaction de l'ensemble de données. La variable 'Amount' est le Montant de la transaction, cette caractéristique peut être utilisée pour l'apprentissage sensible aux coûts dépendant de l'exemple. La fonction 'Class' est la variable de réponse et prend la valeur 1 en cas de fraude et 0 sinon. Compte tenu du rapport de déséquilibre de classes, nous recommandons de mesurer la précision en utilisant l'aire sous la courbe de rappel de précision (AUPRC). La précision de la matrice de confusion n'est pas significative pour une classification non équilibrée.

Tables (réelles)	# d'observations	# de variables	# de classes
VISA	1073	47	2
Fraud-carte-crédit	284 807	31	2

Travail à faire

- 1. Commencer par une étude exploratoire préliminaire
- 2. Utiliser les différentes techniques de classification supervisée vue en cours pour créer un modèle de scoring. Suivant les techniques utilisées (et les fonctions disponibles sous R ou python), vous pourrez utiliser l'ensemble des variables disponibles ou uniquement les variables quantitatives, et réaliser ou non une sélection de variables.
- 3. Comparer l'ensemble de ces techniques à l'aide de courbes ROC (AUC), évaluées soit par validation croisée soit sur échantillon test.

Rapport

Le rapport du projet doit présenter de façon claire et concise:

- l'objet de l'analyse
- la description des données (individus/variables utilisées, variables supplémentaires etc.)
- l'analyse proprement dite
- les commentaires sur les résultats obtenus.

Ce rapport ne devrait pas dépasser 20 pages (les codes sources des programmes utilisés peuvent être mis en annexe). Le projet sera jugé selon les critères suivants:

- Adéquation des méthodes utilisées aux données et problème étudiés.
- Richesse des analyses proposées (au-delà du minimum requis).
- Justesse des commentaires sur les résultats.
- Qualité de la présentation du rapport.

Remise du rapport

Vous devez envoyer votre rapport en format .pdf au plus tard le 6 Janvier 2020 avant minuit à l'adresse suivante <u>l.labiod@gmail.com</u>

Aide. Refaire le traitement proposé dans cet article de blog concernant les imbalanced data (partie avec le package caret) :

https://shiring.github.io/machine_learning/2017/04/02/unbalanced