

REPUBLIQUE DU CAMEROUN  
Paix-Travail-Patrie  
UNIVERSITE DE YAOUNDE 1  
DEPARTEMENT  
D'INFORMATIQUE  
BP/P.O.Box 812  
Yaounde-Cameroun



REPUBLIC OF CAMEROON  
Peace-Work-Fatherland  
UNIVERSITY OF YAOUNDE 1  
COMPUTER SCIENCES  
DEPARTMENT  
BP/P.O.Box 812  
Yaounde-Cameroun

# Prédiction du risque de crédit à base de descripteurs issus de la modélisation des données en graphes

**Noms et prénoms** : Victor Nico DJIEMBOU TIENTCHEU  
**Matricule** : 17T2051  
**Niveau** : Master 2  
**Spécialité** : Sciences de Données (DS)  
**Encadreur** : Dr. Armel Jacques NZEKON NZEKO'O

**Superviseur** : Pr. Maurice TCHUENTE

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Descripteurs extraits des graphes</b>	<b>3</b>
2.1	Généralité et définition . . . . .	3
2.2	Modélisation existant pour extraire les descripteurs des graphes . . . . .	4
<b>3</b>	<b>Descripteurs extraits des graphes multicouches</b>	<b>4</b>
3.1	Processus de construction du graphe . . . . .	5
3.2	Extraction des nouveaux descripteurs du graphe multicouche . . . . .	5
3.3	Prédiction du risque de crédit avec les nouveaux descripteurs . . . . .	7
3.4	Limites et perspectives au travail sur le graphe multicouche . . . . .	7
<b>4</b>	<b>Expérimentation et résultats</b>	<b>8</b>
4.1	Application des descripteurs du graphe multicouche sur le jeu de données AFB (Afriland First Bank) . . . . .	8
4.1.1	Mise en œuvre : intégration des attributs extraits des graphes multicouches dans le processus de prédiction du risque de crédit . . . . .	9
4.2	Application du PageRank personnalisé pour chaque prêt . . . . .	17
4.3	Choix du nombre de couche . . . . .	25
4.3.1	MultiLayer Network (MLN) 1 . . . . .	25
4.3.2	MLN All . . . . .	26
4.3.3	Résultats . . . . .	26
4.4	Choix des attributs à considérer . . . . .	29
4.5	Les meilleures performances . . . . .	29
<b>5</b>	<b>Conclusion et perspectives</b>	<b>30</b>

# 1 Introduction

Les crédits bancaires sont l'une des sources importantes d'enrichissement des banques, cependant, ils peuvent également être source de grandes pertes financières. Afin de minimiser les pertes causées par les activités de prêts, des techniques avancées d'intelligence artificielle sont proposées spécialement pour la prédition du risque de crédit bancaire [?, ?, ?].

La plupart des travaux sur la prédition du risque de crédit considèrent uniquement les attributs de description des prêts (caractéristiques de l'emprunteur et du prêt), et ne s'attardent pas particulièrement sur la modélisation explicite des relations entre les emprunteurs ou même entre les emprunteurs et d'autres caractéristiques considérés dans les données. Ceci peut être une limite, car les individus aux caractéristiques communes peuvent avoir les mêmes comportements de prêts, et donc à partir des comportements connus d'un ensemble d'individus similaires à un individu cible, on peut déduire le comportement de ce dernier.

Dans cette section du projet, il est question de modéliser les données d'une base de prêts bancaires par des graphes dont la définition des noeuds et des arcs est suffisamment pertinente pour que les nouveaux descripteurs extraits de ces graphes contribuent fortement à la décision des modèles de classification supervisée pour la prédition du risque de crédit. L'approche adoptée débute par la construction d'un graphe à partir des données bancaires, se poursuit par l'application des algorithmes de calcul des descripteurs sur des graphes, et se termine par l'ajout des nouveaux descripteurs aux données d'entrée des algorithmes de prédition du risque de crédit. Plus spécifiquement, nous étudions en profondeur l'extraction des descripteurs du graphe multicouches MLN (*MultiLayer Network*) pour la prédition du risque de crédit [1].

La suite du document est structurée comme suit. Nous allons pour ce faire déjà présenter dans la section 2 des généralités sur les graphes et les descripteurs extraits des graphes. Ensuite, nous présenterons dans la section 3 la modélisation en graphes multicouches et les descripteurs extraits de ces graphes. Enfin, nous présenterons quelques expérimentations réalisées et les résultats obtenus à la section 4 et nous achèverons par une conclusion sur le travail fourni et quelques perspectives dans la section 5.

## 2 Descripteurs extraits des graphes

### 2.1 Généralité et définition

Un graphe  $G = \{V, E\}$  est une structure de données qui permet de modéliser les relations entre des entités. La modélisation d'un graphe repose sur deux notions, celle de noeuds  $V$  et celle d'arcs  $E$ . La notion de noeud est associée aux entités qui sont en relation et celle d'arc est associée à la nature de la relation d'une entité (noeud) avec une autre.

Après la construction d'un graphe, il est possible d'extraire des variétés de descripteurs liées soit aux noeuds et donc aux entités, soit à la relation entre les noeuds, et même à la topologie du graphe.

- **Descripteurs d'un noeud :** on peut citer les mesures de centralité qui estiment à quel point le noeud est incontournable dans la navigation dans le graphe. C'est le cas par exemple du PageRank[2] où initialement on attribue le même poids à chaque noeud, puis chaque noeud diffuse son poids à tous ses voisins directs proportionnellement aux poids des relations avec ses voisins. Le processus est répété jusqu'à ce que les poids des noeuds ne changent plus, ou alors jusqu'à ce qu'un nombre maximum d'étapes de diffusion soit atteint. Les noeuds aux poids les plus grands, sont les plus importants.
- **Descripteurs d'une relation entre deux noeuds :** on peut parler des mesures qui décrivent la relation entre deux noeuds à partir du nombre et de la longueur des plus courts chemins entre ces noeuds.

- **Descripteurs de la topologie du graphe :** il est possible d'extraire des communautés dans un graphe (sous-ensemble de noeuds densément connectés entre eux et faiblement connectés au reste du graphe).

Le procédé qui consiste à construire un graphe et à calculer des descripteurs, permet d'apporter de nouvelles informations pour enrichir la description des entités considérées en entrée d'un problème abordé

## 2.2 Modélisation existant pour extraire les descripteurs des graphes

De nombreux modélisation jusqu'ici ont déjà été pensé pour modéliser des graphes avec des base de prêt afin d'extraire des descripteurs pertinents.

S. Mario[3] pour capturer les relations financières entre entités intervenantes (emprunteurs, instituts financiers) dans le contexte d'une inter-coopération des institutions financières et des emprunteur va modéliser un graphes orienté représentant une micro-structure du réseau où chaque noeud représente un emprunteur ou une institution financier et les arêtes, le lien financier existant entre les noeuds. Il va extraire des mesures telles que la taille des k plus proche voisin, le degré d'un noeud, la distance de plus court chemin, le grade d'un noeud, le sous graphe maximal de distance minimale afin d'améliorer l'évaluation du risque de crédit avec des modèles explicables comme Support Vector Machine (SVM) et Logistic Regression (LR).

Xiujuan Xu et Al.[4] vont construire un graphe biparti pour représenter les informations d'historique de prêts d'institut bancaire. Ici chaque noeud matérialise soit un emprunteur soit une modalité caractéristique et une arrête entre un noeud emprunteur et un noeud modalité signifie que l'emprunteur est décrite par cette modalité. Pour prendre en compte les relations complexe entre les emprunteurs, ils vont utiliser trois algorithmes d'analyse de liens basés sur le prétraitement de SVM : Hub Authority Ranking Applicants Algorithm (HARA), Hub-Avg ranking applicants Algorithm (HubAvgRA), Authority-Threshold Algorithm (ATkRA).

Ses deux approches font bien d'utiliser des techniques d'analyse de graphes pour capturer des descripteurs pertinents pouvant améliorer la prédiction du risque de crédit par des algorithmes d'apprentissage automatique, toutefois, elles ne considèrent pas que la relation entre emprunteurs peut être fortement influencée par un sous ensemble de caractéristiques ou niveaux. Hors hypothétiquement, si je suis en relation avec un individu c'est forcement au moins à cause d'un niveau de caractérisation en commun.

## 3 Descripteurs extraits des graphes multicouches

La lecture de l'article « Multilayer network analysis for improved credit risk prediction » de María Óskarsdóttir et Cristián Bravo[1], a été le point d'entrée pour l'usage des descripteurs issus de la modélisation en graphe, pour enrichir les données d'apprentissage des modèles classiques de prédiction du risque de crédit.

En effet, les auteurs utilisent la modélisation par des graphes multicouches (multilayer network) où un emprunteur a autant de noeuds qu'il y a de dimensions qui le caractérisent, et dans chaque dimension, il est relié à des attributs qui le définissent suivant cette dimension. Ainsi, plus les emprunteurs sont similaires, plus ils sont proches dans le graphe multicouche. Une dimension peut par exemple être la localisation géographique ou encore le type d'activité exercé.

Les auteurs s'appuient sur l'idée selon laquelle, les prêts des emprunteurs qui ont un grand nombre de caractéristiques en commun (suivant l'ensemble des dimensions) doivent avoir de grandes probabilités d'être de la même classe.

Ainsi, il se pose les difficultés suivantes :

- Comment établir les relations entre les emprunteurs ?
- Comment déduire des caractéristiques à exploiter à partir de la nouvelle représentation ?
- Comment prédire la classe d'un prêt ?

### 3.1 Processus de construction du graphe

Pour établir les relations entre les emprunteurs, les auteurs construisent un graphe multicouche.

Pour construire un tel graphe, on fixe les dimensions considérées et les attributs associés à chacune de ces dimensions.

Dans le cas de l'article qui s'attarde sur les prêts agricoles, les deux dimensions choisies sont : la localisation géographique et les produits vendus par les agriculteurs.

Les attributs de la dimension localisation géographique peuvent être le district, l'arrondissement etc, et concernant la dimension produit, les attributs peuvent être les différents produits répertoriés.

Dans le graphe multicouche :

- chaque emprunteur a autant de nœuds qu'il y a de couches considérées
- les nœuds de chaque emprunteur sont tous reliés les uns aux autres
- chaque attribut d'une dimension a un nœud associé
- si un emprunteur est décrit par un attribut dans une dimension donnée, alors le nœud emprunteur de cette dimension est relié au nœud attribut associé
- la navigation d'une couche à une autre se fait en passant par les nœuds emprunteurs des différentes couches

### 3.2 Extraction des nouveaux descripteurs du graphe multicouche

Lorsque le graphe multicouche est construit, les nouveaux descripteurs du prêt sont calculés suite à des applications du PageRank Personnalisé sur le graphe résultat.

Les auteurs proposent 03 façons différentes de calculer les nouveaux descripteurs :

- **Intra-influence** : le PageRank Personnalisé est initialisé de manière à favoriser les relations intra-couche dans le processus de diffusion.
- **Inter-influence** : le PageRank Personnalisé est initialisé de manière à favoriser les relations inter-couche dans le processus de diffusion.
- **Influence-combinée** : le PageRank Personnalisé ne favorise pas un type de relation.

Un graphe multicouche M, ayant N nœuds, et L couches, correspond à une représentation de dimension  $N \times N \times L \times L$ , ceci peut être résumé en une matrice carrée  $(N \times L) \times (N \times L)$ .

Dans l'article, les auteurs considèrent deux dimensions pour décrire les emprunteurs dans le graphe multicouche, à savoir la localité et les produits vendus par ces derniers.

Considérons un cas où nous avons 4 emprunteurs (des fermiers), 2 localités (localité des fermiers) et 3 produits (produits agricoles vendus par les fermiers). Dans ce cas de figure, nous avons 3 couches (Emprunteur, Localité et Produit), et nous avons 9 nœuds (4 nœuds emprunteurs + 2 nœuds localités + 3 nœuds produits), et donc la matrice carrée qui permet de représenter le graphe multicouche est de taille  $(9 \times 3) \times (9 \times 3)$

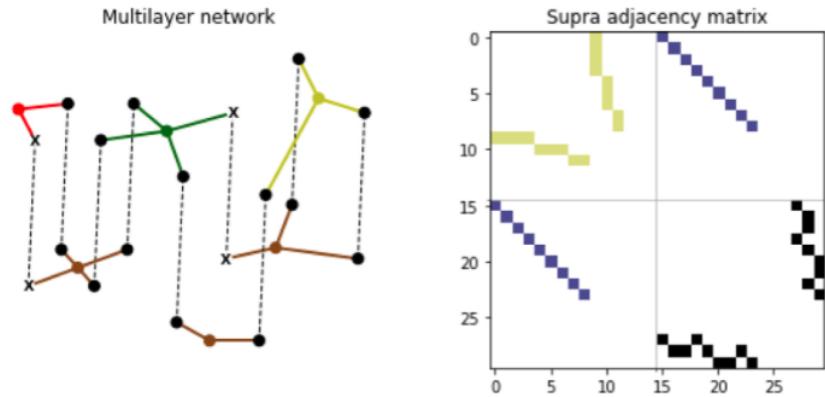


Figure 1 – A gauche on a un exemple de graphe multicouche à deux couches, et à droite on a la représentation de ce graphe sous forme de matrice. Ce graphe contient 9 emprunteurs (nœuds noirs), 3 nœuds de la dimension localité des emprunteurs (nœuds marron) et 3 nœuds de la dimension produits (vert, rouge, jaune). Les relations inter- couches sont matérialisées par des traits interrompus et existent uniquement entre les nœuds emprunteurs qui représentent le même emprunteur dans les différentes couches. Les relations intra-couche sont matérialisées par les autres types de trait (trait marron dans la dimension localité et les autres couleurs dans la dimension produit).

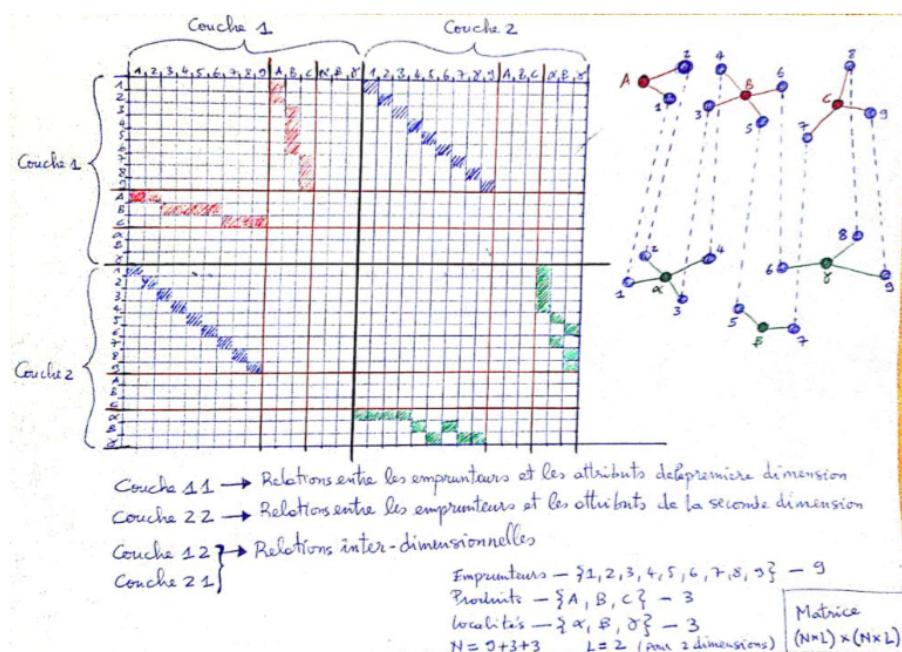


Figure 2 – Représente le même graphe multicouche précédent, mais avec des détails supplémentaires sur le procédé de construction du graphe et de la matrice associée. Un emprunteur (fermier) est relié à sa localité et aux produits agricoles qu'il commercialise.

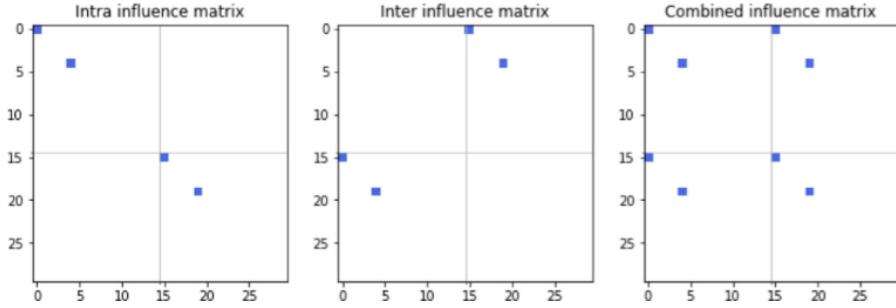


Figure 3 – Les trois scénarios considérés par les auteurs pour calculer les nouveaux descripteurs des prêts par l’application du PageRank Personnalisé sur le graphe multicouche.

Notons que d’autres descripteurs sont considérés dans le graphe multicouche, à l’exemple de :

- Nombre de noeuds emprunteurs qui vendent les mêmes produits que l'emprunteur cible
- Nombre de noeuds emprunteurs défaillants qui vendent les mêmes produits que l'emprunteur cible
- Nombre de noeuds emprunteurs de la même localité que l'emprunteur cible
- Nombre de noeuds emprunteurs défaillants de la même localité que l'emprunteur cible
- Nombre d'emprunteurs de la même localité que l'emprunteur cible et qui vendent les mêmes produits que lui
- Nombre d'emprunteurs défaillants de la même localité que l'emprunteur cible et qui vendent les mêmes produits que lui

### 3.3 Prédiction du risque de crédit avec les nouveaux descripteurs

Pour procéder à la prédiction du risque de crédit avec les modèles classiques d’apprentissage automatique, les descripteurs présents dans le jeu de données, et les nouveaux descripteurs extraits des graphes, sont utilisés comme données d’apprentissage des modèles classiques choisis (Régession logistique et XGBoost) pour la prédiction des risques de crédit.

Une fois que ces modèles sont construits, ces derniers sont utilisés pour prédire les classes des prêts du jeu de test. Dans l’article, les comparaisons des performances des modèles avant et après l’insertion des nouveaux descripteurs, montrent que les nouveaux descripteurs améliorent la qualité des prédictions.

Par ailleurs, les analyses sur l’explicabilité de ces modèles ont montré que les nouveaux descripteurs étaient parmi ceux qui contribuent le plus à la prise de décision des modèles de Régession logistique et XGBoost.

### 3.4 Limites et perspectives au travail sur le graphe multicouche

María Óskarsdóttir et Cristián Bravo[1] dans leur article ont proposé une modélisation des données des prêts en graphe multicouche. Cette approche a pour fort de s’apparenter à la réalité de la vie mais suscite encore des critiques qui définissent en réalité des limites à leur modélisation.

- Tous les attributs ne sont pas considérés dans le graphe multicouche.
- Le choix des attributs catégorielles à considéré comme couche du graphe multicouche est fait de façon arbitraire. Il serait intéressant de proposer un protocole qui permet de faire des choix pertinents.

- Les applications du PageRank Personnalisé sur le graphe multicouche ne sont pas suffisamment personnalisées, car le niveau de personnalisation reste à l'échelle intra et inter collectif (pour tous les nœuds concernés). Cependant, on pourrait avoir 03 applications différentes du PageRank Personnalisé pour chacun des prêts. Dans cette optique, seuls les nœuds qui sont liés au prêt courant peuvent recevoir la valeur 1.

Ses critiques nous permettent d'émettre ses possibles extensions :

- Exploiter un graphe qui prend en compte tous les attributs descriptifs des prêts. On peut par exemple avoir un nœud pour chacune des modalités possibles de chaque attribut. Ensuite, relier tous les nœuds du même prêt ou alors incrémenter les poids des arcs qui relient tous les nœuds des modalités d'un prêt. Ensuite, appliquer le PageRank Personnalisé par chaque prêt, sur le graphe résultat, afin de ressortir avec de nouveaux descripteurs du prêt à l'exemple de l'estimer la classe de ce prêt par le PageRank.
- Proposer un protocole qui permet de choisir efficacement les attributs catégoriels à considérer dans le graphe multicouche.
- Personnaliser davantage les exécutions du PageRank Personnalisé de manière à affecter la valeur 1, uniquement aux nœuds associés au prêt courant pour lequel on calcule les valeurs des descripteurs.
- Faire ce travail sur plusieurs jeux de données

## 4 Expérimentation et résultats

### 4.1 Application des descripteurs du graphe multicouche sur le jeu de données AFB (Afriland First Bank)

Pour expérimenter les concepts appris, il est nécessaire de choisir des attributs qualitatifs qui vont représenter les dimensions (couches) du graphe multicouche. Nous avons donc recensé les attributs catégoriels du jeu de données AFB de Afriland First Bank.

Lorsqu'on ignore la classe des prêts, les attributs catégoriels de ce jeu de données sont :

- **Type / Motif** : le type de prêt ou motif du prêt bancaire
- **Fonction** : le métier ou l'occupation de l'emprunteur
- **Civilité** : civilité de l'emprunteur (Monsieur, Madame, Mademoiselle)
- **Statut matrimonial** : statut matrimonial de l'emprunteur (Célibataire, Marié, Divorcé)

Après avoir recensé les attributs catégoriels, nous avons choisi d'implémenter trois graphes multicouches à deux couches. Le premier graphe multicouche nommé ici **MLN1** est construit à partir des attributs **Fonction & Civilité**, le second graphe **MLN2** est construit à partir des attributs **Fonction & Statut-Matrimonial** et enfin, le troisième graphe **MLN3** est construit à partir des attributs **Fonction & Motif**.

Les attributs extraits de ces différents graphes multicouches sont énumérés comme suit pour une paire d'attribut (Att1, Att2) :

- **MLN\_Att1\_degré** : le nombre d'emprunteur avec la même valeur d'Att
- **MLN\_Att2\_degré** : le nombre d'emprunteur avec la même valeur d'Att
- **MLN\_Att1\_et\_Att2\_degré** : le nombre d'emprunteur avec la même valeur Att1 et Att2
- **MLN\_bipart\_intra\_Att1\_Att2** : le score PageRank maximal entre le noeud de l'emprunt de couche Att1 et Att2 lorsque seul les noeuds intra (modalités Att1 et Att2) sont inclus dans le vecteur de personnalisation du PageRank

- **MLN\_bipart\_inter\_Att1\_Att2** : le score PageRank maximal entre le noeud de l'emprunt de couche Att1 et Att2 lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN\_bipart\_combine\_Att1\_Att2** : le score PageRank maximal entre le noeud de l'emprunt de couche Att1 et Att2
- **MLN\_bipart\_intra\_Att1\_max** : le score PageRank maximal du noeud de Att1 associé à un emprunt lorsque seul les noeuds intra (modalités de Att1) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN\_bipart\_inter\_Att1\_max** : le score PageRank maximal de Att1 associé à un emprunt lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN\_bipart\_combine\_Att1\_max** : le score PageRank maximal du noeud de Att1 associé à un emprunt
- **MLN\_bipart\_intra\_Att2\_max** : le score PageRank maximal du noeud de Att2 associé à un emprunt lorsque seul les noeuds intra (modalités de Att2) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN\_bipart\_inter\_Att2\_max** : le score PageRank maximal du noeud de Att2 associé à un emprunt lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN\_bipart\_combine\_Att2\_max** : le score PageRank maximal du noeud de Att2 associé à un emprunt

#### 4.1.1 Mise en œuvre : intégration des attributs extraits des graphes multicouches dans le processus de prédiction du risque de crédit

Nous avons considéré cinq modèles classiques de l'apprentissage automatique pour la prédiction du risque de crédit : Decision Tree (DT), Random Forest (RF), LR, eXtreme Gradient Boosting (XGB) et SVM.

Pour chacun des graphes multicouches considérés (MLN1, MLN2, et MLN3), chaque modèle de prédiction est appliqué 04 fois. Chacune des applications du modèle diffère de l'autre par l'ensemble d'attributs descripteurs considérés :

- **Classic** : les attributs considérés sont tous ceux fournis avec le jeu de données
- **Classic + MLN** : on considère tous les attributs du jeu de données et on ajoute les autres attributs extraits du graphe multicouches MLN
- **Classic – Att** : on considère une partie des attributs fournis avec le jeu de données. Ceux qui sont liés aux dimensions du graphe multicouches MLN sont ignorés. Par exemple, pour le cas MLN1, les attributs relatifs à Fonction et à Civilités seront complètement écartés de la phase d'apprentissage
- **Classic + MLN - Att** : on écarte les attributs relatifs aux dimensions du graphe multicouche MLN, et on intègre les attributs extraits du graphe multicouche MLN

Nous pouvons ainsi évaluer l'impact des attributs choisis dans le graphe multicouches associés à leur représentation standard fourni dans le jeu de données (**Classic + MLN**), sans leur représentation standard (**Classic + MLN - Att**), et enfin évaluer l'impact de leur absence des données d'apprentissage des modèles (**Classic – Att**).

Le tableau ci-dessous présente l'ensemble des résultats obtenus pour les cinq différents modèles de prédiction SVM, XGBoost, Arbre de décision, Régression Logistique et Forêt Aléatoire, avec les trois graphes multicouches considérés (MLN1, MLN2, MLN3) et suivant les métriques Exactitude, Précision, Rappel et F1-score

		Accuracy	Precision	Recall	F1-score
SVM	classic	<b>0.9577</b>	0.9518	<b>0.9577</b>	<b>0.9577</b>
	classic + MLN1	0.9576	0.9517	0.9576	0.9576
	classic + MLN2	<b>0.9577</b>	<b>0.952</b>	<b>0.9577</b>	<b>0.9577</b>
	classic + MLN3	0.9576	0.9515	0.9576	0.9576
	classic - AMLN1	<b>0.9577</b>	0.9516	<b>0.9577</b>	<b>0.9577</b>
	classic - AMLN2	<b>0.9577</b>	0.9516	<b>0.9577</b>	<b>0.9577</b>
	classic - AMLN3	0.8153	0.9348	0.8154	0.8118
	classic - AMLN1 + MLN1	0.9575	0.9515	0.9575	0.9575
	classic - AMLN2 + MLN2	0.9575	0.9513	0.9575	0.9575
	classic - AMLN3 + MLN3	0.9553	0.9481	0.9553	0.9553
XGBOOST	classic	0.9954	0.9961	0.9954	0.9954
	classic + MLN1	<b>0.9962</b>	<b>0.9977</b>	<b>0.9962</b>	<b>0.9962</b>
	classic + MLN2	0.9956	0.997	0.9956	0.9956
	classic + MLN3	0.9961	0.997	0.9961	0.9961
	classic - AMLN1	0.9947	0.9954	0.9947	0.9947
	classic - AMLN2	0.9952	0.9961	0.9952	0.9952
	classic - AMLN3	0.9558	0.9813	0.9558	0.9558
	classic - AMLN1 + MLN1	0.9953	0.9959	0.9953	0.9953
	classic - AMLN2 + MLN2	0.9947	0.9956	0.9947	0.9947
	classic - AMLN3 + MLN3	0.9959	0.997	0.9959	0.9959
DECISION TREE	classic	0.9959	<b>0.9988</b>	0.9959	0.9959
	classic + MLN1	0.9945	0.9984	0.9945	0.9945
	classic + MLN2	0.9948	0.9984	0.9948	0.9948
	classic + MLN3	0.9932	0.9986	0.9932	0.9932
	classic - AMLN1	0.9947	0.9986	0.9947	0.9947
	classic - AMLN2	0.9948	<b>0.9988</b>	0.9948	0.9948
	classic - AMLN3	0.968	0.9871	0.968	0.9679
	classic - AMLN1 + MLN1	0.9949	0.9986	0.9949	0.9949
	classic - AMLN2 + MLN2	<b>0.9961</b>	0.9986	<b>0.9961</b>	<b>0.9961</b>
	classic - AMLN3 + MLN3	0.9949	0.9984	0.9949	0.9949
LOGISTIC REGRESSION	classic	0.9588	0.9638	0.9588	0.9588
	classic + MLN1	0.959	0.9638	0.959	0.959
	classic + MLN2	0.959	0.964	0.959	0.959
	classic + MLN3	0.9584	0.9636	0.9584	0.9584
	classic - AMLN1	0.9584	<b>0.9649</b>	0.9584	0.9584
	classic - AMLN2	0.9584	<b>0.9649</b>	0.9584	0.9584
	classic - AMLN3	0.8163	0.9394	0.8164	0.8126
	classic - AMLN1 + MLN1	0.9591	<b>0.9649</b>	0.9591	0.9591
	classic - AMLN2 + MLN2	0.9582	0.9646	0.9582	0.9582
	classic - AMLN3 + MLN3	<b>0.9593</b>	0.9602	<b>0.9593</b>	<b>0.9593</b>
RANDOM FOREST	classic	0.996	0.9993	0.996	0.996
	classic + MLN1	0.9952	0.9991	0.9952	0.9952
	classic + MLN2	0.9958	0.9993	0.9958	0.9958
	classic + MLN3	0.9951	0.9993	0.9951	0.9951
	classic - AMLN1	<b>0.9963</b>	0.9993	<b>0.9963</b>	<b>0.9963</b>
	classic - AMLN2	0.9962	<b>0.9995</b>	0.9962	0.9962
	classic - AMLN3	0.966	0.9918	0.966	0.966
	classic - AMLN1 + MLN1	0.9955	0.9991	0.9955	0.9955
	classic - AMLN2 + MLN2	0.9959	0.9993	0.9959	0.9959
	classic - AMLN3 + MLN3	0.9959	<b>0.9995</b>	0.9959	0.9959

Tableau 1 – Récapitulatif des résultats avec les différents graphes multicouches

Lorsqu'on s'attarde sur le classement des modèles de prédiction, le modèle de prédiction associé à la meilleure performance est la Forêt Aléatoire (Random Forest), suivi d'Arbre de décision et de XGB. La Régression Logistique et SVM ferment ce classement.

En s'intéressant aux meilleures performances obtenues avec les attributs extraits des graphes multicouches, on constate que le graphe multicouche MLN1 est le meilleur pour le modèle XGBoost, le graphe MLN2 est meilleur pour SVM et Arbre de décision et enfin le graphe MLN3 est le meilleur pour Régression Logistique et Forêt Aléatoire.

Si on considère uniquement les cas de figure où chaque modèle est associé au graphe multicouche qui lui correspond le mieux, on fait les remarques suivantes :

- **SVM** : il ne faut pas considérer les attributs issus du graphe multicouche
- **XGBoost** : les meilleures performances sont atteintes lorsqu'on considère à la fois les attributs extraits du graphe multicouche ainsi que leur forme classique dans le jeu de données. Et la forme classique de ces attributs a plus d'impact que les attributs extraits du graphe multicouche.
- **Arbre de décision** : les meilleures performances sont atteintes lorsqu'on considère uniquement les attributs issus du graphe multicouche et qu'on ignore ces attributs dans leur

représentation classique.

- **Régression Logistique** : les meilleures performances sont atteintes lorsqu'on considère uniquement les attributs issus du graphe multicouches et qu'on ignore ces attributs dans leur représentation classique.
- **Forêt Aléatoire** : en considérant la Précision comme métrique d'évaluation, la meilleure performance est atteinte lorsqu'on considère uniquement les attributs issus du graphe multicouches et qu'on ignore ces attributs dans leur représentation classique.

D'après les résultats obtenus, la considération des attributs extraits des graphes multicouches (Classic + MLNi et Classic – AMLNi + MLNi) permet l'amélioration des performances des modèles Arbre de décision, XGBoost, Random Forest et Régression Logistique. L'unique modèle pour lequel il n'y a pas d'amélioration mais un statu quo c'est le modèle SVM.

#### Evaluation de la contribution des attributs pour les modèles de prédiction

Suite à l'appréciation des résultats des modèles de prédiction, nous nous sommes intéressés aux contributions des différents attributs pour l'obtention des résultats considérés, notamment pour les cas de figure associés aux meilleurs graphes multicouches. Les graphiques des cinq pages suivantes illustrent ces informations (une page par modèle de prédiction). Sur chaque graphique, plus une barre est longue et orientée vers la droite, plus l'attribut associé à cette barre contribue positivement aux décisions du modèle de prédiction. Par contre plus une barre est étirée vers la gauche, plus l'attribut contribue négativement aux décisions du modèle de prédiction. Sur les graphiques, les barres peuvent avoir trois couleurs possibles :

- **Les barres bleus** ; correspondent aux attributs classiques fournis dans le jeu de données, mais qui sont différents des attributs considérés pour le graphe multicouches.
- **Les barres jaunes** : correspondent aux attributs classiques fournis dans le jeu de données et qui sont considérés pour la construction du graphe multicouches.
- **Les barres vertes** : correspondent aux nouveaux descripteurs issus du graphe multicouches correspondant.

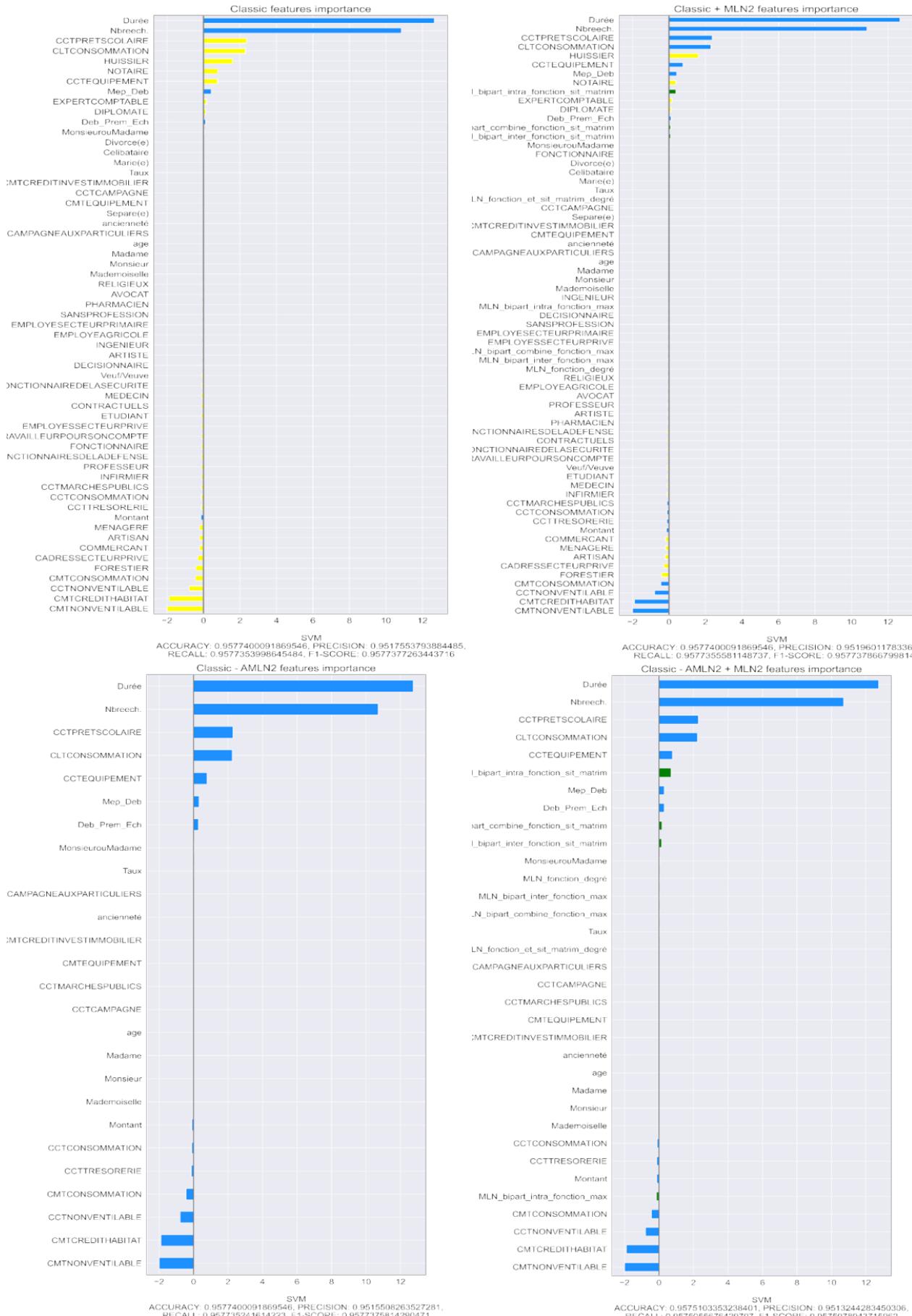
En observant les graphiques des résultats des combinaisons de type Classic + MLNi, on constate que pour les modèles de prédiction

- Forêt Aléatoire (MLN\_bipart\_intra\_Motif\_max)
- XGboost (MLN\_Fonction\_et\_Civilité\_Degrée)
- et Arbre de Décision (MLN\_bipart\_intra\_Fonction\_Sit\_Matrim)

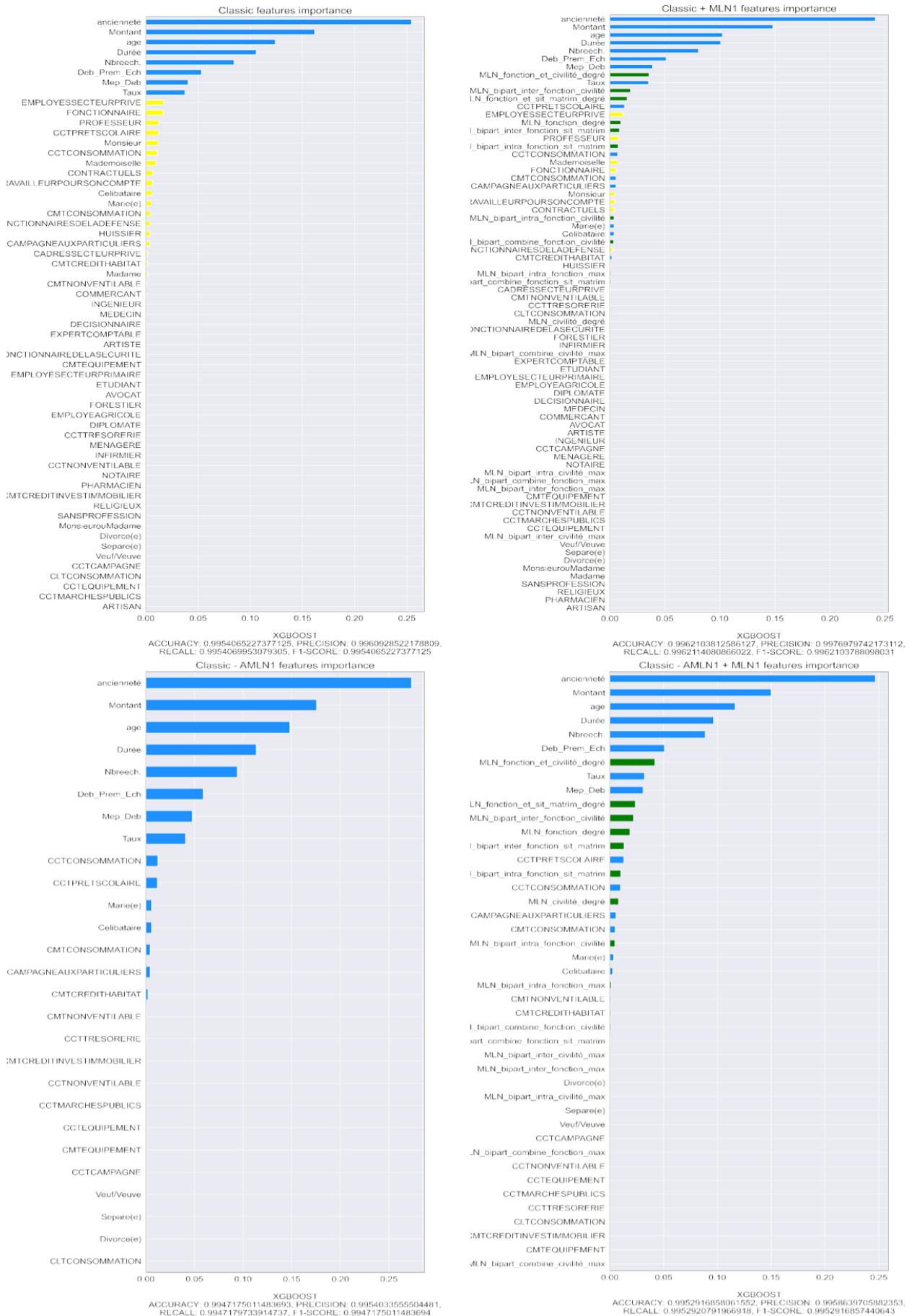
les descripteurs issus du graphe multicouches sont mieux classés que les attributs associés sous leur formes classique

provenant du jeu de données. Par contre, la tendance est inversée pour les cas de SVM et la Régression Logistique où les descripteurs classiques sont mieux classés que ceux extraits des graphes multicouche. Ce constat renforce la pertinence des descripteurs extraits des graphes multicouches, car ces derniers contribuent beaucoup plus pour les modèles de prédiction associés aux plus grandes performances. Ce qui est davantage renforcé par le cas de la Forêt Aléatoire qui est le modèle le plus performant de tous, et dont l'attribut MLN\_bipart\_intra\_Motif\_max extrait des graphes multicouches est celui qui contribue le plus aux prises de décisions.

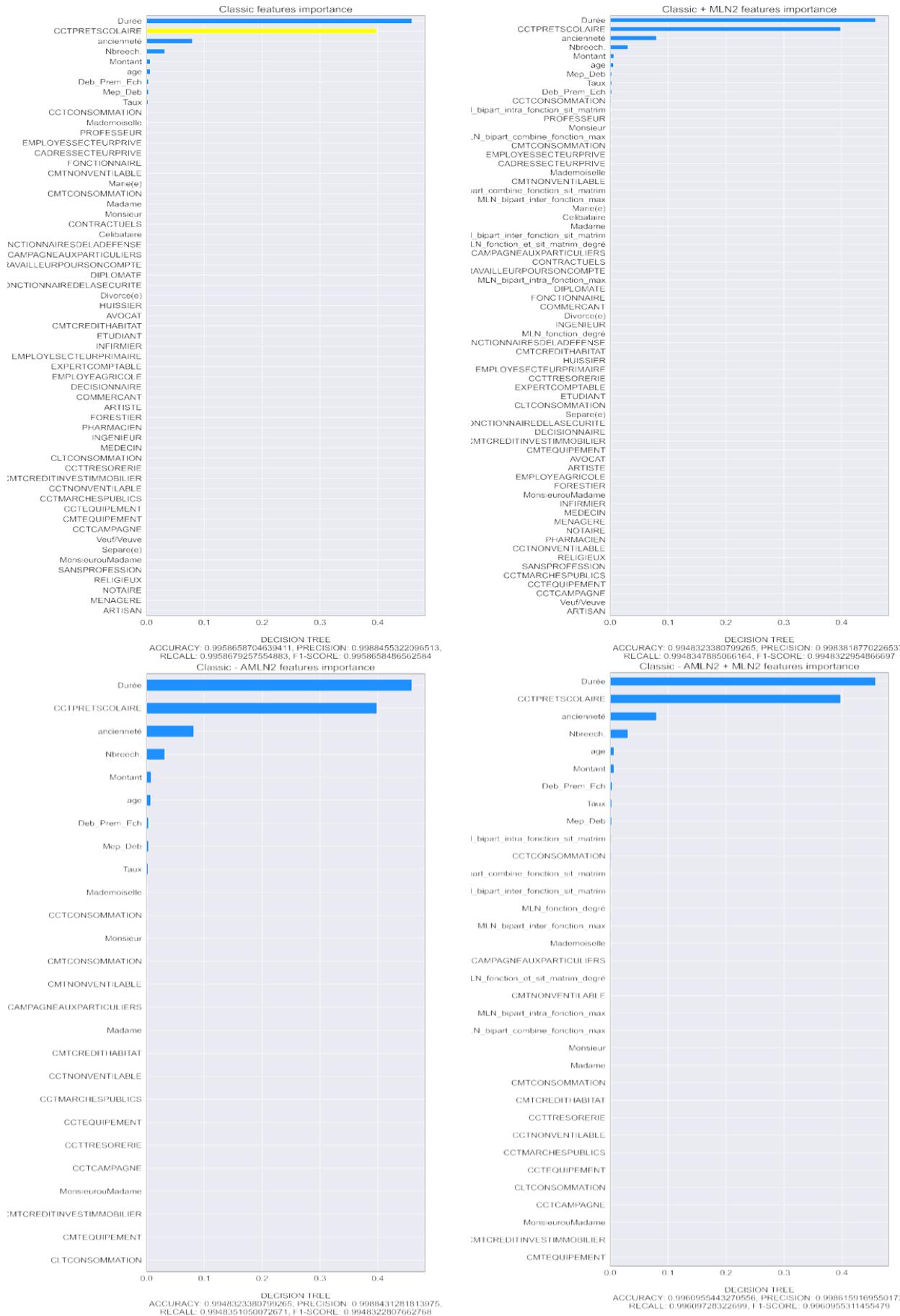
- Cas du modèle SVM – meilleur graphe multicouche MLN2



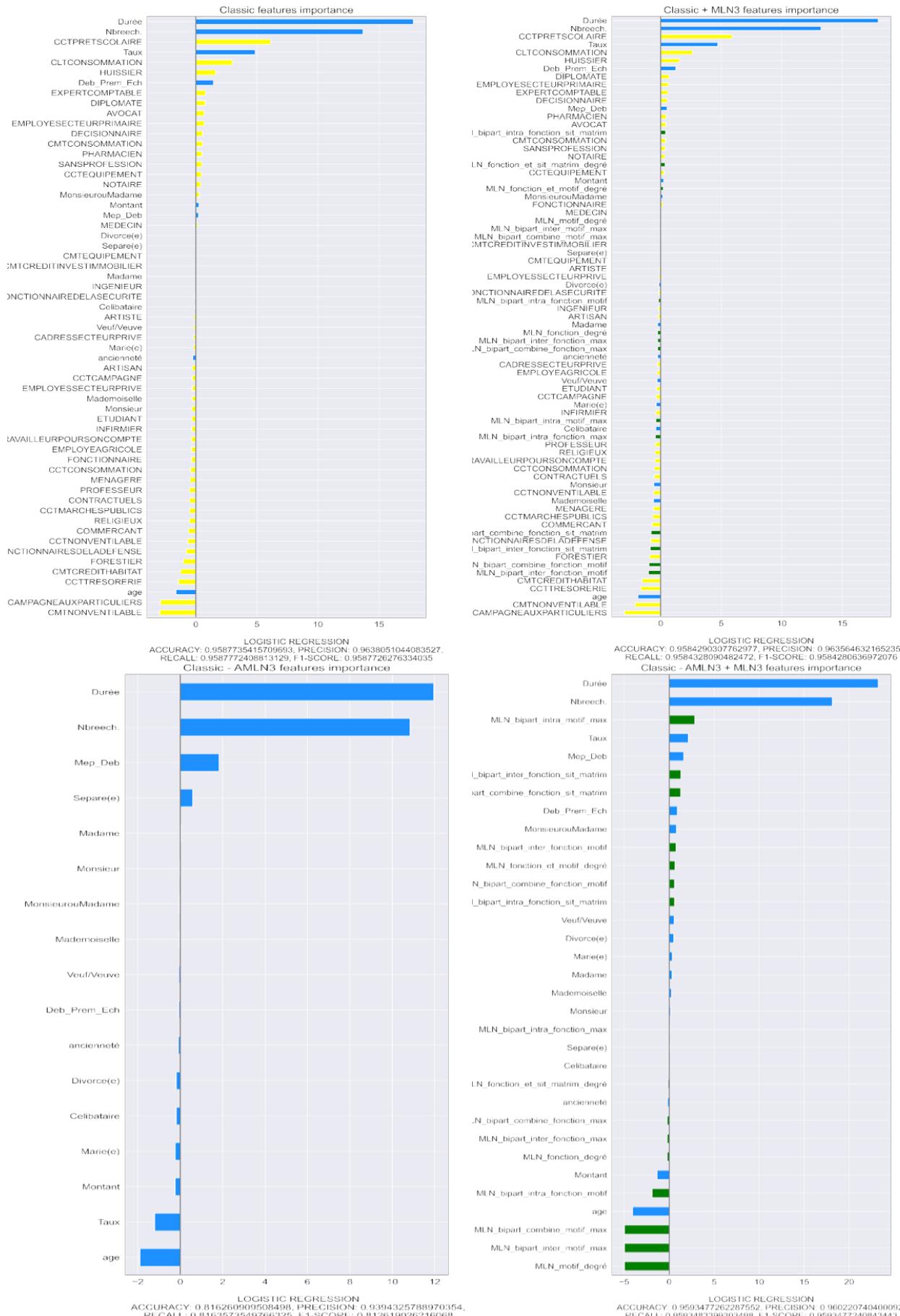
- Cas du modèle XGBoost – meilleur graphe multicouche MLN1



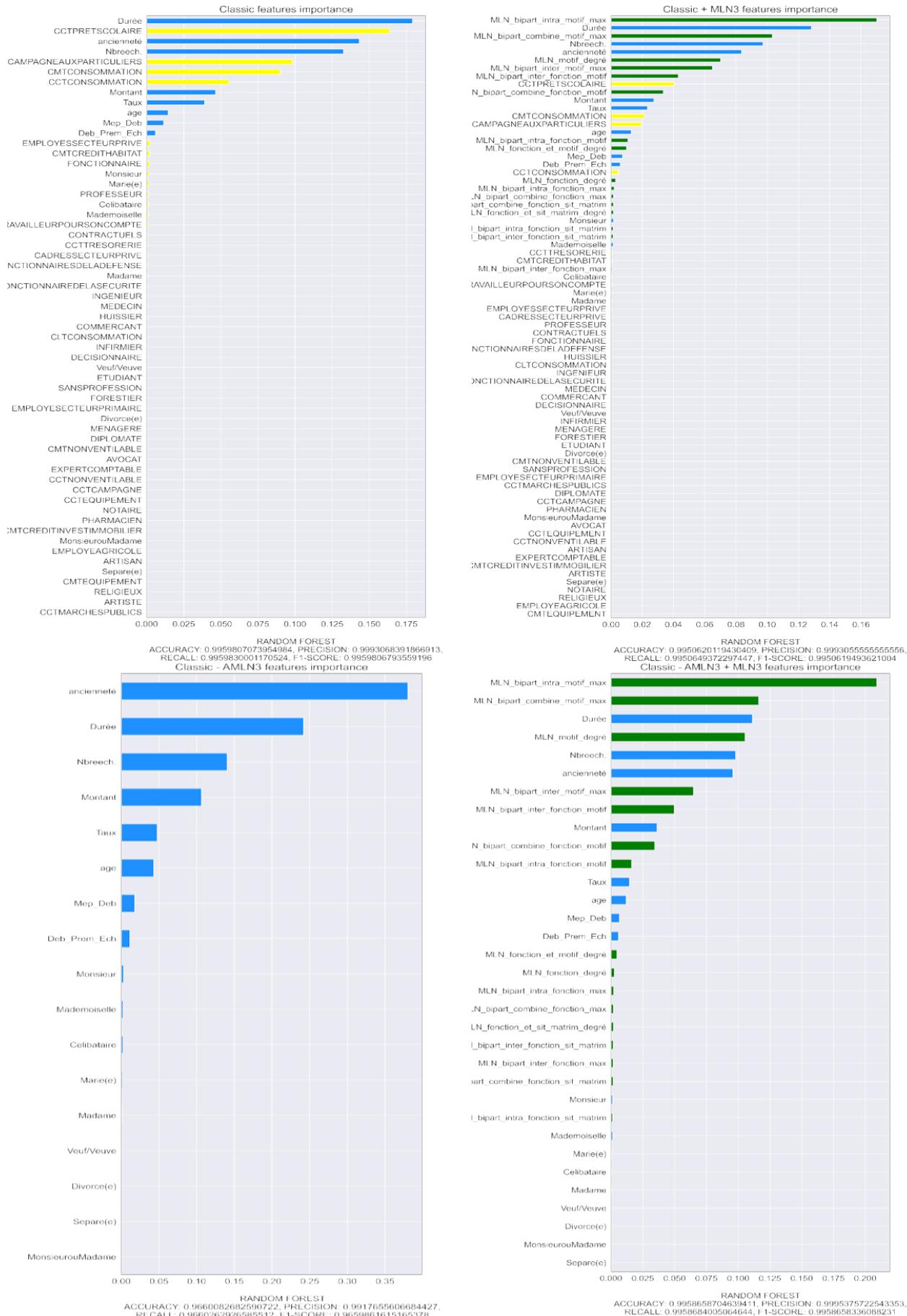
- Cas du modèle Arbre de Décision (Decision Tree) – meilleur graphe multicouche MLN2



- Cas du modèle Régression Logistique – meilleur graphe multicouche MLN3



- Cas du modèle Forêt Aléatoire (Random Forest) – meilleur graphe multicouche MLN3



## 4.2 Application du PageRank personnalisé pour chaque prêt

Nous proposons dans cette expérimentation qu'un score de PageRank soit calculé pour chaque emprunts de tel sorte que pour un emprunt précis, l'exécution du PageRank permettant d'extraire ce score n'aille comme unique noeuds possible de démarrage initiale ceux associés à cet emprunt. Nous espérons par cette ultra personnalisation de l'extraction d'influence dans le réseau formé proposer des descripteurs pertinentes dans l'évaluation du risque de crédit.

L'équation de calcul itératif du PageRank personnalisé à un seul emprunteur est donnée par :

$$PR_{i+1}^U = (1 - \alpha) \times d + \alpha \times PR_i^U \times H, \quad (1)$$

$$\text{sc. } |d| = L \times N + M, \quad \sum_{i=0}^{|d|} d[i] = 1, \quad d[i] = 0 \iff d[i] \notin U \quad (2)$$

où :

$PR_i^U$  est le PageRank à l'itération i,

$PR_{i+1}^U$  est le PageRank à l'itération i+1,

$\alpha$  est le facteur d'amortissement (typiquement  $\alpha = 0,85$ ),

$N$  est le nombre total d'emprunts dans le graphe

$H$  est la matrice de transition du graphe .

$d$  est le vecteur de probabilités de démarrage

$L$  est le nombre de couches

$N$  est le nombre d'emprunts dans le graphe

$M$  est la somme du nombre de modalités d'attributs représenté par chaque couche

$U$  est l'emprunteur courant.

AER		Accuracy			Precision			Recall			F1-score		
		Classic	G <sup>^</sup>	P <sup>^</sup>	Classic	G <sup>^</sup>	P <sup>^</sup>	Classic	G <sup>^</sup>	P <sup>^</sup>	Classic	G <sup>^</sup>	P <sup>^</sup>
xgb	Classic + MLN	0.9697	0.0	0.0	0.9949	0.0	0.0	0.9746	0.0	0.0	0.9588	0.0	0.0
	Classic + MLN - Att	0.9697	0.0	0.0	0.9949	0.0	0.0	0.9746	0.0	0.0	0.9588	0.0	0.0
dtc	Classic + MLN	0.9735	0.0	0.0	0.9949	0.0	0.0	0.977	0.0	0.0	0.9637	0.0	0.0
	Classic + MLN - Att	0.9735	0.0	0.0	0.9949	0.0	0.0	0.977	0.0	0.0	0.9637	0.0	0.0
lrc	Classic + MLN	0.8977	0.0	0.0	0.9583	0.0	0.0	0.8876	0.0	0.0	0.8643	0.0	0.0
	Classic + MLN - Att	0.8977	<b>2.95</b>	<b>2.95</b>	0.9583	<b>1.65</b>	<b>1.65</b>	0.8876	<b>3.89</b>	<b>3.89</b>	0.8643	<b>4.01</b>	<b>4.01</b>
rfc	Classic + MLN	0.9735	0.0	0.0	1.0	0.0	0.0	0.9828	0.0	0.0	0.9641	0.0	0.0
	Classic + MLN - Att	0.9735	0.0	0.0	1.0	0.0	0.0	0.9828	0.0	0.0	0.9641	0.0	0.0
sv	Classic + MLN	0.8902	0.0	0.0	0.9833	0.0	0.0	0.9114	0.0	0.0	0.8621	0.0	0.0
	Classic + MLN - Att	0.8902	1.27	<b>1.7</b>	0.9833	<b>1.7</b>	<b>1.7</b>	0.9114	2.7	<b>2.96</b>	0.8621	1.84	<b>2.33</b>
Total			1	<b>2</b>		<b>2</b>	<b>2</b>		1	<b>2</b>		1	<b>2</b>

Tableau 2 – Pourcentage (%) de gain PageRank avec personnalisation global versus PageRank avec personnalisation à un seul emprunteur pour le jeu de données AER de Kaggle

AFB		Accuracy			Precision			Recall			F1-score		
		Classic	G^	P^	Classic	G^	P^	Classic	G^	P^	Classic	G^	P^
xgb	Classic + MLN	0.9891	0.12	<b>0.16</b>	0.9915	0.09	<b>0.18</b>	0.9842	0.17	<b>0.29</b>	0.9854	0.17	<b>0.21</b>
	Classic + MLN - Att	0.9891	0.07	<b>0.12</b>	0.9915	0.07	<b>0.09</b>	0.9842	0.12	<b>0.17</b>	0.9854	0.1	<b>0.17</b>
dtc	Classic + MLN	0.9841	<b>0.05</b>	0.02	0.9901	0.04	0.0	0.9795	<b>0.06</b>	-0.04	0.9788	<b>0.07</b>	0.02
	Classic + MLN - Att	0.9841	0.02	<b>0.04</b>	0.9901	0.02	<b>0.04</b>	0.9795	0.01	<b>0.02</b>	0.9788	0.02	<b>0.05</b>
lrc	Classic + MLN	0.9636	0.0	0.0	0.9832	0.0	0.0	0.9591	0.0	0.0	0.9521	0.0	0.0
	Classic + MLN - Att	0.9636	-0.04	0.0	0.9832	-0.03	0.0	0.9591	-0.05	0.0	0.9521	-0.05	0.0
rfc	Classic + MLN	0.986	<b>0.12</b>	0.07	0.9892	0.0	<b>0.05</b>	0.9798	<b>0.08</b>	<b>0.08</b>	0.9813	<b>0.16</b>	0.09
	Classic + MLN - Att	0.986	<b>0.12</b>	<b>0.12</b>	0.9892	<b>0.09</b>	0.07	0.9798	<b>0.18</b>	0.14	0.9813	<b>0.16</b>	<b>0.16</b>
sv	Classic + MLN	0.9613	<b>0.02</b>	<b>0.02</b>	0.9833	<b>0.01</b>	<b>0.01</b>	0.9578	<b>0.01</b>	<b>0.01</b>	0.9493	<b>0.02</b>	<b>0.02</b>
	Classic + MLN - Att	0.9613	0.0	0.0	0.9833	0.0	0.0	0.9578	0.0	0.0	0.9493	0.0	0.0
Total			4	5		2	5		4	5		4	5

Tableau 3 – Pourcentage (%) de gain PageRank avec personnalisation global versus PageRank avec personnalisation à un seul emprunteur pour le jeu de données AFB

CREDIT_RISK_DATASET		Accuracy			Precision			Recall			F1-score		
		Classic	G^	P^	Classic	G^	P^	Classic	G^	P^	Classic	G^	P^
xgb	Classic + MLN	0.9263	<b>0.08</b>	<b>0.08</b>	0.938	0.12	<b>0.27</b>	0.847	0.17	<b>0.18</b>	0.8804	0.12	<b>0.14</b>
	Classic + MLN - Att	0.9263	0.12	<b>0.17</b>	0.938	0.0	<b>0.18</b>	0.847	<b>0.48</b>	0.45	0.8804	0.27	<b>0.33</b>
dtc	Classic + MLN	0.8825	<b>0.1</b>	-0.06	0.7289	<b>0.34</b>	0.23	0.827	0.13	<b>0.25</b>	0.827	<b>0.15</b>	-0.07
	Classic + MLN - Att	0.8825	-0.22	<b>0.24</b>	0.7289	-0.88	<b>0.82</b>	0.827	-0.1	<b>0.37</b>	0.827	-0.25	<b>0.35</b>
lrc	Classic + MLN	0.8489	0.01	<b>0.04</b>	0.7311	0.07	<b>0.08</b>	0.7152	0.04	<b>0.1</b>	0.7434	0.04	<b>0.09</b>
	Classic + MLN - Att	0.8489	-0.02	0.0	0.7311	<b>0.71</b>	0.7	0.7152	-0.06	0.0	0.7434	-0.05	0.0
rfc	Classic + MLN	0.9162	<b>0.35</b>	0.22	0.9296	<b>0.33</b>	0.06	0.825	<b>0.84</b>	0.64	0.8616	<b>0.7</b>	0.48
	Classic + MLN - Att	0.9162	0.19	<b>0.35</b>	0.9296	0.09	<b>0.98</b>	0.825	0.55	<b>0.64</b>	0.8616	0.37	<b>0.63</b>
sv	Classic + MLN	0.8502	0.02	0.0	0.7783	0.0	<b>0.1</b>	0.6992	<b>0.11</b>	0.04	0.7329	<b>0.11</b>	0.01
	Classic + MLN - Att	0.8502	0.0	0.0	0.7783	0.0	<b>0.69</b>	0.6992	0.0	0.0	0.7329	0.0	0.0
Total			3	5		3	7		3	5		3	5

Tableau 4 – Pourcentage (%) de gain PageRank avec personnalisation global versus PageRank avec personnalisation à un seul emprunteur pour le jeu de données CREDIT RISK

GERMAN		Accuracy			Precision			Recall			F1-score		
		Classic	G^	P^	Classic	G^	P^	Classic	G^	P^	Classic	G^	P^
xgb	Classic + MLN	0.745	<b>7.38</b>	4.7	0.7021	13.94	<b>15.3</b>	0.6819	<b>9.59</b>	5.88	0.6919	<b>10.12</b>	6.3
	Classic + MLN - Att	0.745	6.04	<b>6.71</b>	0.7021	14.56	<b>15.3</b>	0.6819	7.41	<b>8.46</b>	0.6919	7.72	<b>8.73</b>
dtc	Classic + MLN	0.62	<b>10.48</b>	<b>10.48</b>	0.45	<b>24.29</b>	23.02	0.5659	14.47	<b>14.67</b>	0.567	13.99	<b>14.69</b>
	Classic + MLN - Att	0.62	10.48	<b>11.29</b>	0.45	<b>24.29</b>	24.18	0.5659	13.5	<b>15.92</b>	0.567	13.79	<b>15.87</b>
lrc	Classic + MLN	0.745	0.67	0.0	0.7111	2.28	0.0	0.6786	0.56	0.0	0.6888	0.65	0.0
	Classic + MLN - Att	0.745	2.01	<b>4.03</b>	0.7111	8.99	<b>18.66</b>	0.6786	2.18	<b>4.36</b>	0.6888	2.44	<b>4.89</b>
rfc	Classic + MLN	0.73	<b>4.79</b>	2.74	0.75	<b>13.13</b>	7.53	0.6407	<b>6.77</b>	4.96	0.6447	<b>8.03</b>	5.93
	Classic + MLN - Att	0.73	<b>6.85</b>	4.11	0.75	12.28	<b>12.51</b>	0.6407	<b>10.11</b>	5.48	0.6447	<b>11.83</b>	6.48
sv	Classic + MLN	<b>0.755</b>	<b>1.32</b>	0.66	0.7561	<b>5.13</b>	2.5	0.683	<b>1.61</b>	0.56	0.6945	<b>1.8</b>	0.66
	Classic + MLN - Att	0.755	1.99	<b>3.31</b>	0.7561	6.55	<b>7.23</b>	0.683	3.13	<b>4.26</b>	0.6945	3.35	<b>4.71</b>
Total			5	5		4	5		4	5		4	5

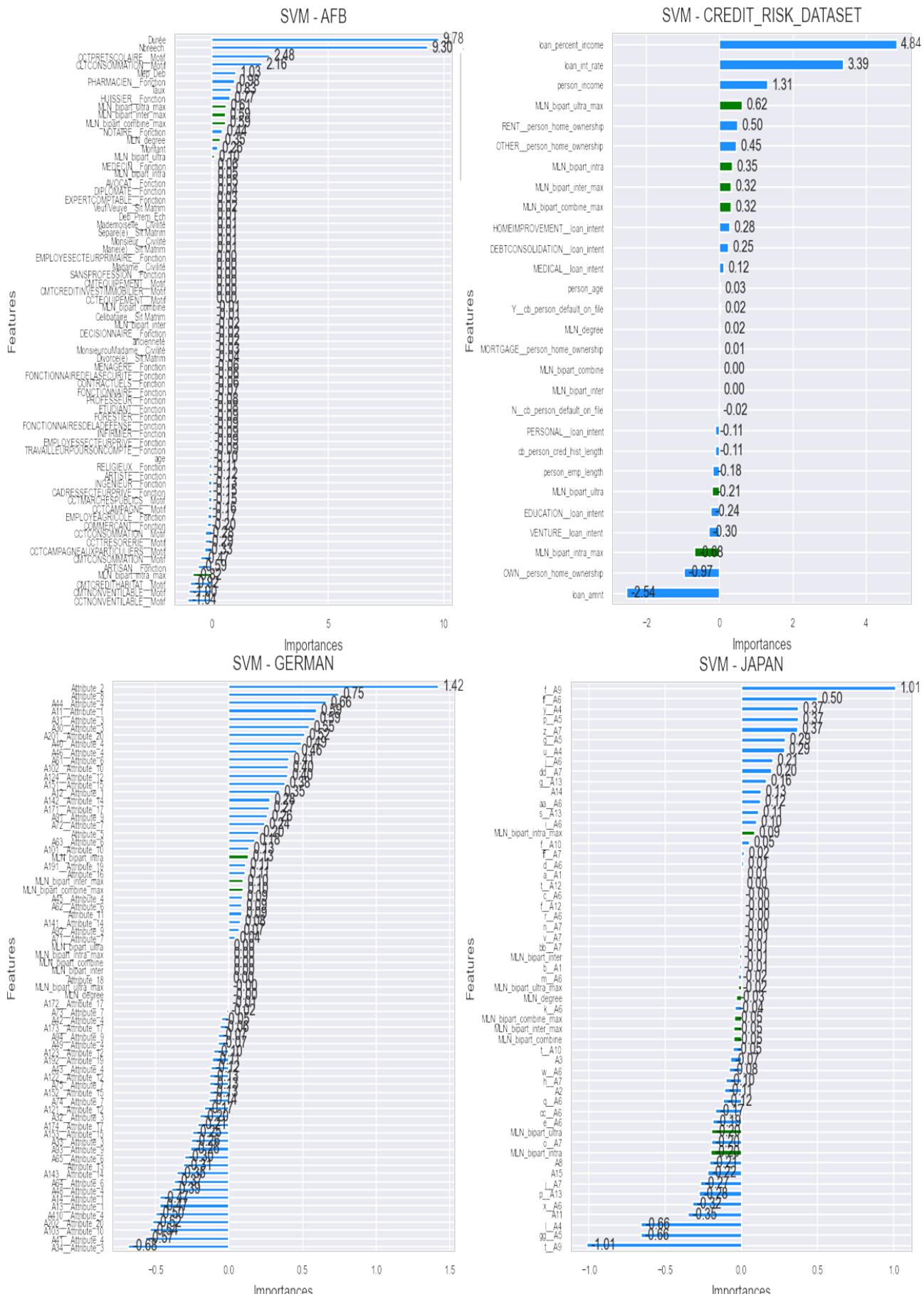
Tableau 4 – Pourcentage (%) de gain PageRank avec personnalisation global versus PageRank avec personnalisation à un seul emprunteur pour le jeu de données GERMAN

JAPAN		Accuracy			Precision			Recall			F1-score		
		Classic	G <sup>^</sup>	P <sup>^</sup>	Classic	G <sup>^</sup>	P <sup>^</sup>	Classic	G <sup>^</sup>	P <sup>^</sup>	Classic	G <sup>^</sup>	P <sup>^</sup>
xgb	Classic + MLN	0.8551	3.39	<b>5.09</b>	0.913	1.82	<b>2.03</b>	0.8599	3.22	<b>4.73</b>	0.8546	3.37	<b>5.05</b>
	Classic + MLN - Att	0.8551	<b>5.09</b>	3.39	0.913	<b>2.03</b>	1.82	0.8599	<b>4.73</b>	3.22	0.8546	<b>5.05</b>	3.37
dtc	Classic + MLN	0.8116	3.57	<b>4.46</b>	0.84	4.37	<b>5.44</b>	0.8107	3.84	<b>4.22</b>	0.8096	3.68	<b>4.43</b>
	Classic + MLN - Att	0.8116	<b>6.25</b>	5.36	0.84	<b>6.18</b>	5.44	0.8107	<b>6.45</b>	5.02	0.8096	<b>6.36</b>	5.31
lrc	Classic + MLN	0.8551	0.0	0.0	0.913	0.0	0.0	0.8599	0.0	0.0	0.8546	0.0	0.0
	Classic + MLN - Att	0.8551	0.0	0.0	0.913	0.0	<b>1.36</b>	0.8599	0.0	<b>0.2</b>	0.8546	0.0	<b>0.02</b>
rfc	Classic + MLN	0.8768	<b>0.83</b>	<b>0.83</b>	0.9167	1.3	<b>1.41</b>	0.8794	0.55	<b>0.93</b>	0.876	0.79	<b>0.84</b>
	Classic + MLN - Att	0.8768	<b>0.83</b>	<b>0.83</b>	0.9167	<b>0.12</b>	<b>0.12</b>	0.8794	<b>0.74</b>	<b>0.74</b>	0.876	<b>0.82</b>	<b>0.82</b>
sv	Classic + MLN	0.8478	0.0	0.0	0.9242	0.0	0.0	0.8551	0.0	0.0	0.8476	0.0	0.0
	Classic + MLN - Att	0.8478	<b>0.86</b>	<b>0.86</b>	0.9242	0.0	<b>1.55</b>	0.8551	0.36	<b>0.96</b>	0.8476	0.79	<b>0.87</b>
Total			<b>5</b>	<b>5</b>		3	<b>6</b>		3	<b>6</b>		3	<b>6</b>

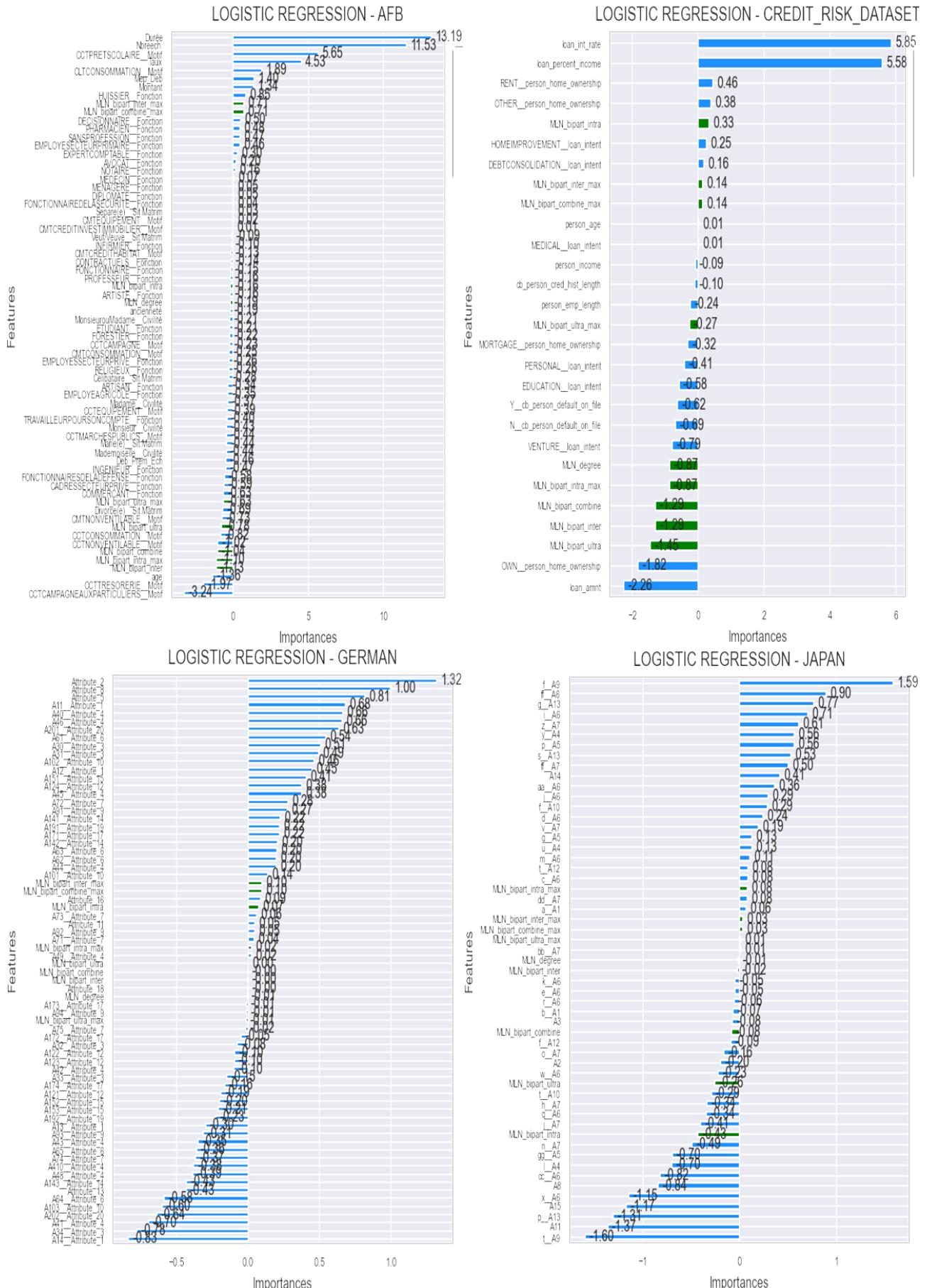
Tableau 5 – Pourcentage (%) de gain PageRank avec personnalisation global versus PageRank avec personnalisation à un seul emprunteur pour le jeu de données JAPAN

Nous remarquons que dans les résultats ci-haut que bien qu'il y ai une sorte d'équivalence entre la personnalisation globale et celle à un seul emprunteur, bien de fois, la version personnalisation à un seul emprunteur l'emporte sur globale et nous battons ainsi le meilleur modèle issu de ces jeux de données. De plus, nous pouvons aussi visualiser les résultats d'analyse d'importance des attributs dans la décision suivant selon les modèles :

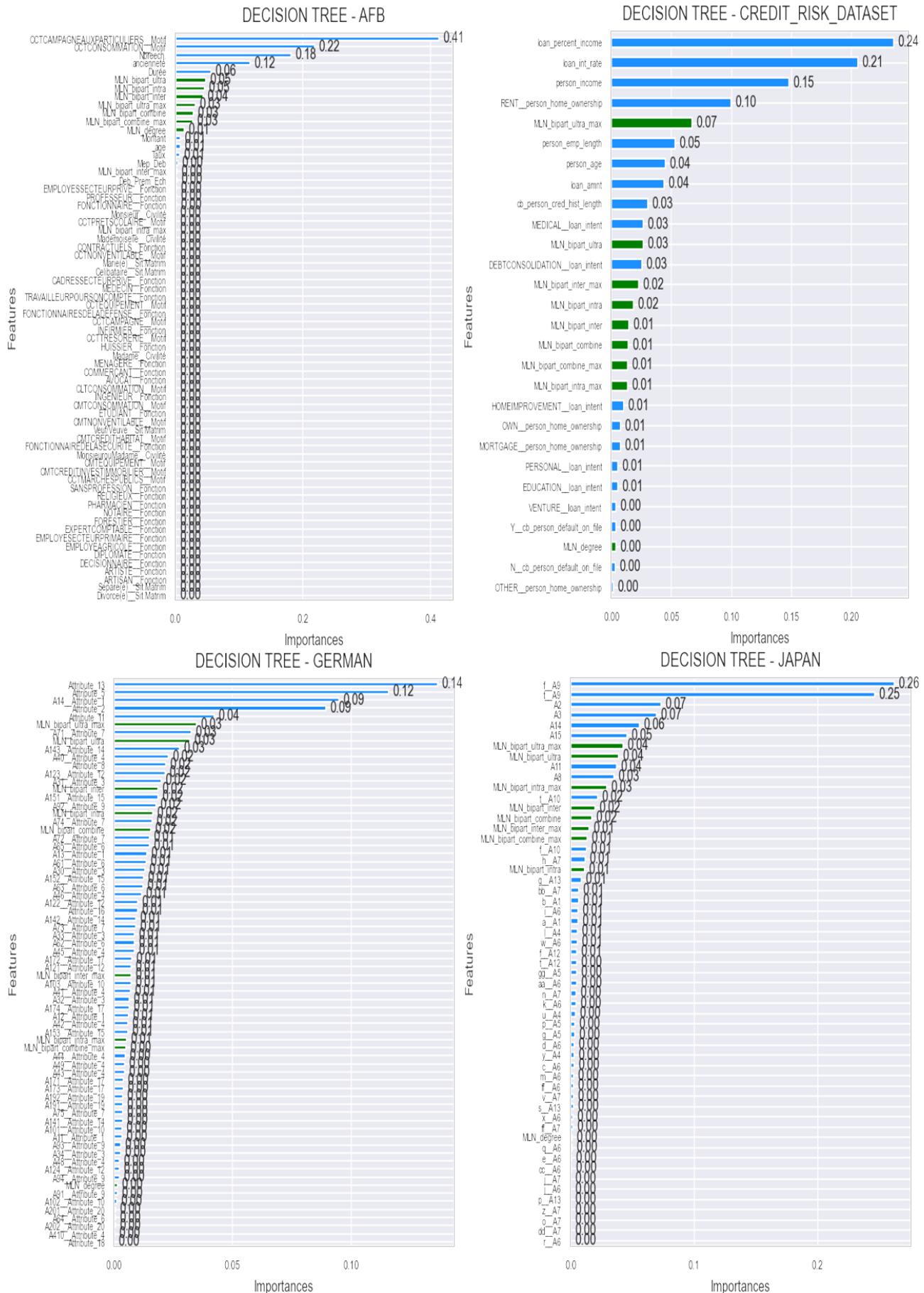
- SVM



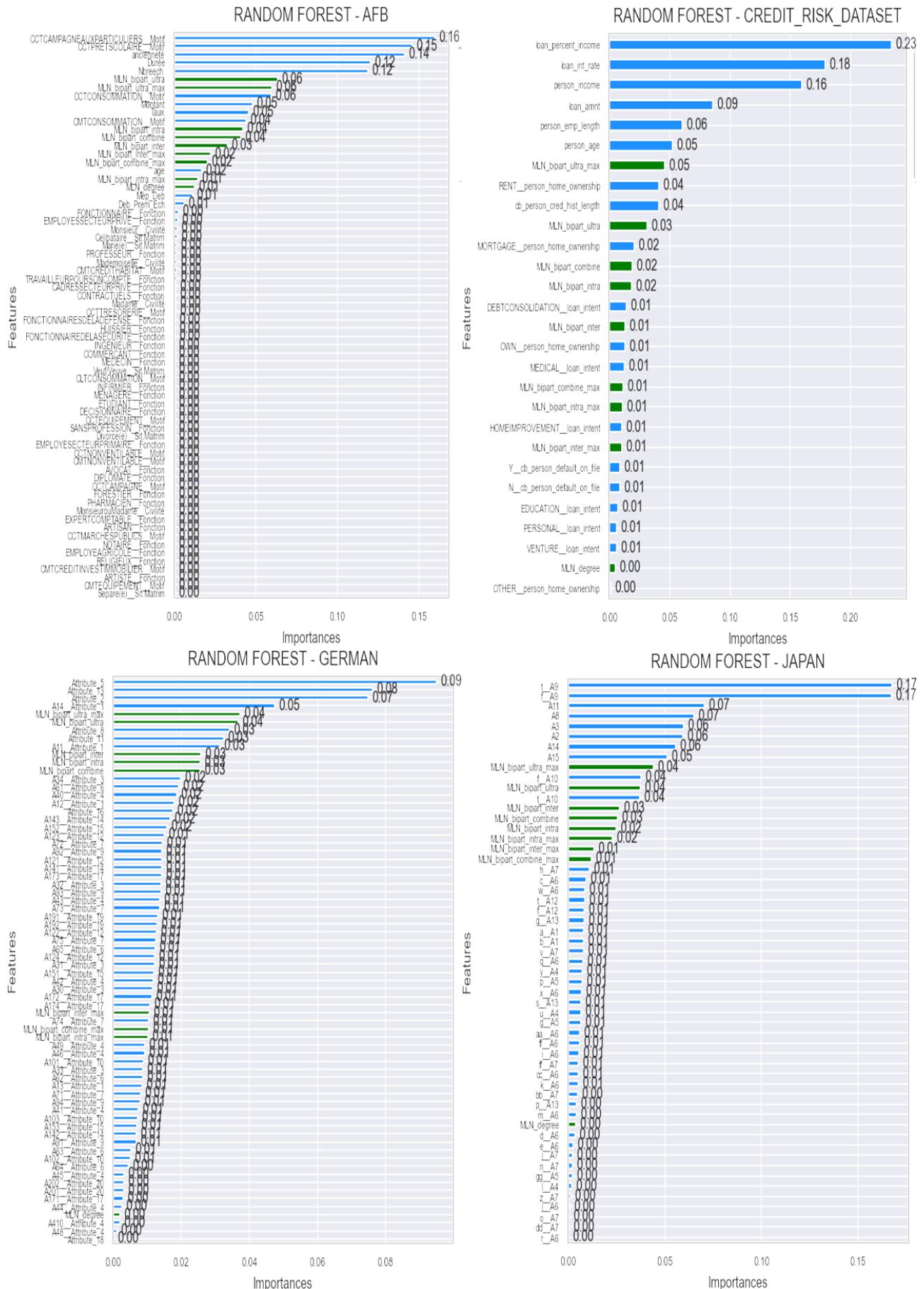
- LR



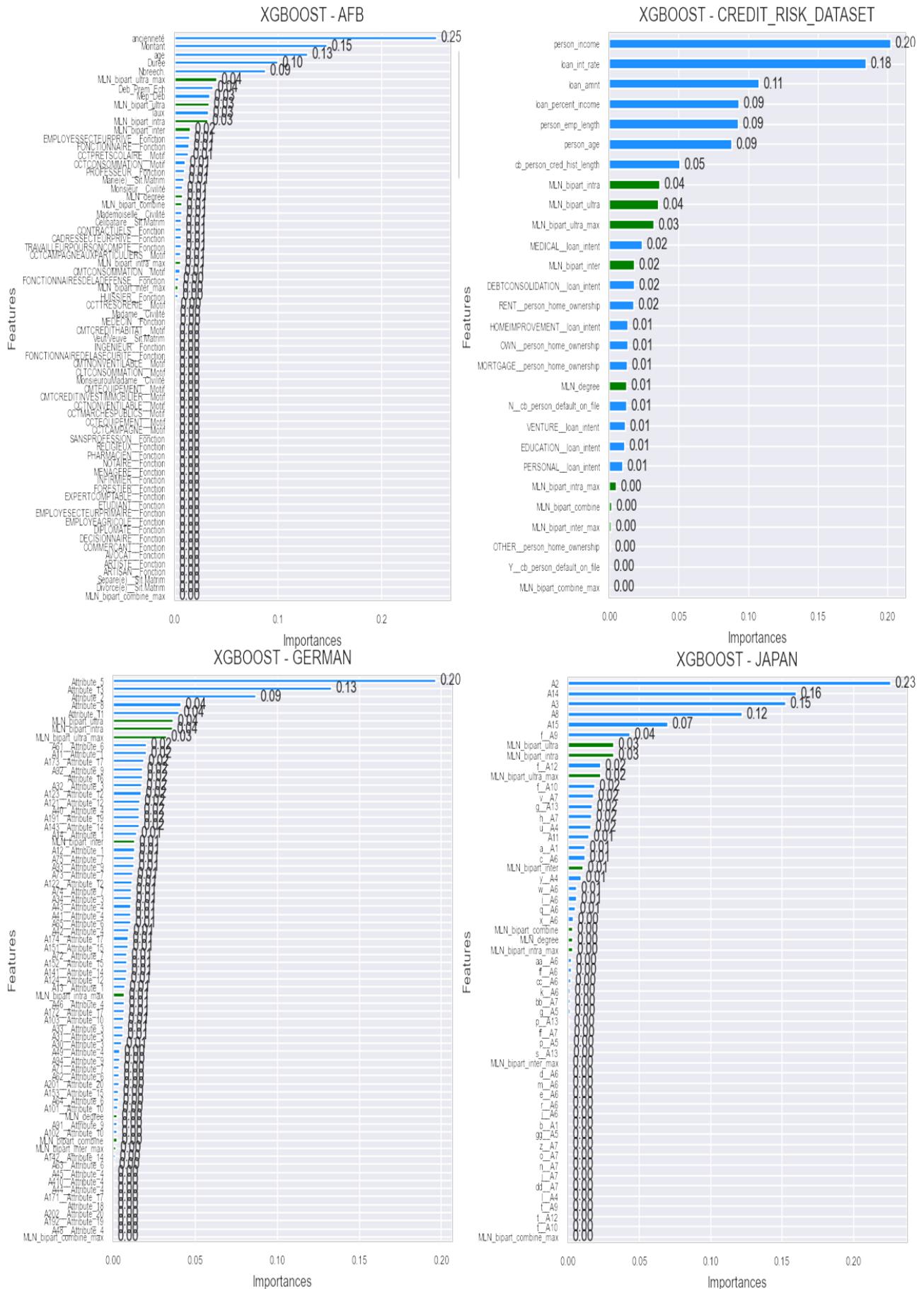
- DT



- RF



- XGB



Observations :

- AER : Nous remarquons de part ses graphiques dans le contexte du jeu de données AER, les descripteurs issus des graphes ont une tendance à ne pas forte expliquer autant les décision de possibilité défaut de prêt ou de solvabilité. De plus la meilleure explication est obtenu à partir du modèle SVM
- AFB : Pour ce qui concerne AFB, les descripteurs issus des graphes font partir des attributs qui expliquent le mieux la décision des modèles. Encore plus loin, nous remarquons que les descripteurs résultant du PageRank personnalisé à un seul emprunt surpassé la version de l'article de base.
- CREDIT RISK : Ces résultats présentent les mêmes caractéristiques que la AFB, et une fois de plus, nous avons la supériorité des descripteurs du PageRank avec personnalisation à un emprunt.
- GERMAN : Les modèles basé sur des arbres pour ce jeux de données sont ceux pour qui les attributs des graphes expliquent le mieux les sortir en particulier la version issus de la personnalisation à un seul emprunt du PageRank
- JAPAN : Les attributs des graphes de façon généraliste permettent d'expliquer les sorties pour ce jeu de données. Et une fois de plus, PageRank personnalisé à un seul emprunt surpassé la version à personnalisation globale de l'article.

En somme, nous remarquons de part ses graphiques que les descripteurs issus des graphes en moyenne pour chaque jeu de données se trouve dans le top 10. Cependant nous devons mettre une emphase sur les descripteurs ultra personnalisé qui eux en générale se trouve dans le top 5 pour chaque modèle d'apprentissage que se soit en faveur de la classe positive ou négative. Ceci dis que pourrait apporter le choix d'une meilleur configuration du MLN en terme de pertinence de descripteurs extraits et de métriques de classification ?

### 4.3 Choix du nombre de couche

Pour expérimenter l'impact du nombreux de couches sur la qualité des descripteurs choisis, nous avons construire un graphe multicouche à 1 seule couche ou un seul attribut catégoriel (MLN 1) et un autre graphe multicouche avec autant de couche qu'il y'a d'attributs de type dans l'ensemble de données (MLN All).

#### 4.3.1 MLN 1

Comme mise en hypothèse, il peut s'avérer que les relations entre les emprunts soient essentiellement porter sur un seul niveau, caractéristique ou couche. Alors, pour vérifier cela, nous avons modélisé nos données sous forme de graphes multicouche biparti à une seule couche.

Pour chaque attribut catégoriel du jeu de données, nous construisons ces graphes et extraions les informations suivante où **case\_k** représente le nom de l'attribut qui sert à construire le MLN 1 :

- **MLN\_case\_k\_degré** : le nombre d'emprunteur avec la même case\_k
- **MLN\_bipart\_intra\_case\_k** : le score PageRank du noeud de l'emprunt de couche case\_k lorsque seul les noeuds intra (modalités motif) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN\_bipart\_inter\_case\_k** : le score PageRank du noeud de l'emprunt de couche case\_k lorsque seul les noeuds intra (noeuds emprunt) sont inclus dans le vecteur de personnalisation du PageRank

- **MLN\_bipart\_combine\_case\_k** : le score PageRank du noeud de l'emprunt de couche case\_k
- **MLN\_bipart\_intra\_max\_case\_k** : le score PageRank du noeud de la modalité de couche case\_k lorsque seul les noeuds intra (modalités motif) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN\_bipart\_inter\_max\_case\_k** : le score PageRank du noeud de la modalité de couche case\_k lorsque seul les noeuds intra (noeuds emprunt) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN\_bipart\_combine\_max\_case\_k** : le score PageRank du noeud de la modalité de couche case\_k
- **MLN\_bipart\_ultra\_case\_k** : le score PageRank du noeud de l'emprunt de couche case\_k lorsque seul les noeuds associés à l'emprunt (un seul emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN\_bipart\_ultra\_max\_case\_k** : le score PageRank du noeud de la modalité de couche case\_k lorsque seul les noeuds associés à l'emprunt (un seul emprunteur) sont inclus dans le vecteur de personnalisation du PageRank

#### 4.3.2 MLN All

Nous expérimentons dans cette partie, l'impact d'une modélisation des emprunts avec autant de couches qu'il en a de données catégorielles sur l'évaluation du risque de crédit financier. Chaque couche dans cette modélisation met en relation chaque emprunt à sa modalité de la caractéristique de la couche et les différentes couches communiquent entre elles par le biais des noeuds matérialisant l'emprunt. Comme dans MLN 1, nous avons extraire les mêmes descripteurs.

#### 4.3.3 Résultats

AER		Accuracy			Precision			Recall			F1-score			
		classic	MLN 1^	MLN All^	classic	MLN 1^	MLN All^	classic	MLN 1^	MLN All^	classic	MLN 1^	MLN All^	
xgb	Classic + MLN	Global	0.9697	0.0	0.0	0.9949	0.0	0.0	0.9746	0.0	0.0	0.9588	0.0	0.0
		Personalized	0.9697	0.0	0.0	0.9949	0.0	0.0	0.9746	0.0	0.0	0.9588	0.0	0.0
	Classic + MLN - Att	Global	0.9697	0.0	0.0	0.9949	0.0	0.0	0.9746	0.0	0.0	0.9588	0.0	0.0
		Personalized	0.9697	0.0	0.0	0.9949	0.0	0.0	0.9746	0.0	0.0	0.9588	0.0	0.0
dtc	Classic + MLN	Global	0.9735	0.0	-0.39	0.9949	0.0	-0.5	0.977	0.0	-0.84	0.9637	0.0	-0.56
		Personalized	0.9735	0.0	0.0	0.9949	0.0	0.0	0.977	0.0	0.0	0.9637	0.0	0.0
	Classic + MLN - Att	Global	0.9735	0.0	0.0	0.9949	0.0	0.0	0.977	0.0	0.0	0.9637	0.0	0.0
		Personalized	0.9735	-0.39	0.0	0.9949	-0.5	0.0	0.977	-0.84	0.0	0.9637	-0.56	0.0
lrc	Classic + MLN	Global	0.8977	0.0	0.0	0.9583	0.0	0.0	0.8876	0.0	0.0	0.8643	0.0	0.0
		Personalized	0.8977	0.0	0.0	0.9583	0.0	0.0	0.8876	0.0	0.0	0.8643	0.0	0.0
	Classic + MLN - Att	Global	0.8977	0.0	<b>2.95</b>	0.9583	0.0	<b>1.65</b>	0.8876	0.0	<b>3.89</b>	0.8643	0.0	<b>4.01</b>
		Personalized	0.8977	2.12	<b>2.95</b>	0.9583	1.11	<b>1.65</b>	0.8876	2.68	<b>3.89</b>	0.8643	2.85	<b>4.01</b>
rfc	Classic + MLN	Global	0.9735	0.0	0.0	1.0	0.0	0.0	0.9828	0.0	0.0	0.9641	0.0	0.0
		Personalized	0.9735	0.0	0.0	1.0	0.0	0.0	0.9828	0.0	0.0	0.9641	0.0	0.0
	Classic + MLN - Att	Global	0.9735	0.0	0.0	1.0	0.0	0.0	0.9828	0.0	0.0	0.9641	0.0	0.0
		Personalized	0.9735	0.0	0.0	1.0	0.0	0.0	0.9828	0.0	0.0	0.9641	0.0	0.0
sv	Classic + MLN	Global	0.8902	0.0	0.0	0.9833	0.0	0.0	0.9114	0.0	0.0	0.8621	0.0	0.0
		Personalized	0.8902	0.0	0.0	0.9833	0.0	0.0	0.9114	0.0	0.0	0.8621	0.0	0.0
	Classic + MLN - Att	Global	0.8902	0.84	<b>1.27</b>	0.9833	0.57	<b>1.7</b>	0.9114	1.16	<b>2.7</b>	0.8621	1.1	<b>1.84</b>
		Personalized	0.8902	<b>1.7</b>	<b>1.7</b>	0.9833	<b>1.7</b>	<b>1.7</b>	0.9114	<b>2.96</b>	<b>2.96</b>	0.8621	<b>2.33</b>	<b>2.33</b>
Total				1	<b>4</b>		1	<b>4</b>		1	<b>4</b>		1	<b>4</b>

Tableau 6 – Gain d'impact des modélisations pour le jeu de données AER de Kaggle

AFB			Accuracy				Precision				Recall				F1-score			
			classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^
xgb	Classic + MLN	Global	0.9891	<b>0.12</b>	0.11	0.07	0.9915	<b>0.09</b>	<b>0.09</b>	0.02	0.9842	<b>0.17</b>	0.16	0.07	0.9854	<b>0.17</b>	0.14	0.1
		Personalized	0.9891	0.05	<b>0.16</b>	0.04	0.9915	0.07	<b>0.18</b>	0.07	0.9842	0.09	<b>0.29</b>	0.09	0.9854	0.07	<b>0.21</b>	0.05
	Classic + MLN - Att	Global	0.9891	<b>0.07</b>	<b>0.07</b>	-0.15	0.9915	<b>0.07</b>	<b>0.07</b>	-0.16	0.9842	0.09	<b>0.12</b>	-0.26	0.9854	<b>0.1</b>	<b>0.1</b>	-0.21
		Personalized	0.9891	<b>0.12</b>	0.05	-0.36	0.9915	<b>0.09</b>	0.05	-0.25	0.9842	<b>0.17</b>	0.08	-0.51	0.9854	<b>0.17</b>	0.07	-0.49
dtc	Classic + MLN	Global	0.9841	0.0	<b>0.05</b>	-0.12	0.9901	-0.02	<b>0.04</b>	-0.07	0.9795	-0.05	<b>0.06</b>	-0.15	0.9788	0.0	<b>0.07</b>	-0.16
		Personalized	0.9841	-0.07	<b>0.02</b>	-0.14	0.9901	-0.05	0.0	-0.09	0.9795	-0.09	<b>-0.04</b>	-0.19	0.9788	-0.09	<b>0.02</b>	-0.18
	Classic + MLN - Att	Global	0.9841	<b>0.02</b>	0.0	-0.35	0.9901	0.0	<b>0.02</b>	-0.32	0.9795	<b>0.01</b>	-0.05	-0.57	0.9788	<b>0.02</b>	0.0	-0.47
		Personalized	0.9841	<b>0.04</b>	-0.03	-0.66	0.9901	0.0	<b>0.04</b>	-0.46	0.9795	<b>0.02</b>	0.01	-0.92	0.9788	<b>0.05</b>	-0.05	-0.89
lrc	Classic + MLN	Global	0.9636	-0.06	0.0	-0.06	0.9832	0.0	0.0	-0.03	0.9591	-0.04	0.0	-0.06	0.9521	-0.07	0.0	-0.07
		Personalized	0.9636	0.0	-0.04	-0.06	0.9832	0.0	0.0	-0.05	0.9591	0.0	-0.02	-0.08	0.9521	0.0	-0.04	-0.07
	Classic + MLN - Att	Global	0.9636	<b>-0.04</b>	<b>-0.04</b>	-16.26	0.9832	<b>-0.03</b>	-0.05	-11.9	0.9591	<b>-0.05</b>	-0.07	-23.36	0.9521	<b>-0.05</b>	<b>-0.05</b>	-22.43
		Personalized	0.9636	0.0	-0.11	-16.17	0.9832	0.0	-0.07	-11.91	0.9591	0.0	-0.15	-23.32	0.9521	0.0	-0.15	-22.36
rfc	Classic + MLN	Global	0.986	0.02	<b>0.12</b>	-0.03	0.9892	-0.11	0.0	-0.11	0.9798	-0.1	<b>0.08</b>	-0.14	0.9813	0.02	<b>0.16</b>	-0.05
		Personalized	0.986	<b>0.07</b>	0.05	-0.14	0.9892	0.03	<b>0.05</b>	-0.18	0.9798	0.07	<b>0.08</b>	-0.28	0.9813	<b>0.09</b>	0.07	-0.19
	Classic + MLN - Att	Global	0.986	0.05	<b>0.12</b>	-0.26	0.9892	-0.04	<b>0.09</b>	-0.25	0.9798	-0.01	<b>0.18</b>	-0.43	0.9813	0.06	<b>0.16</b>	-0.36
		Personalized	0.986	<b>0.12</b>	0.1	-0.51	0.9892	0.05	<b>0.07</b>	-0.23	0.9798	0.13	<b>0.14</b>	-0.57	0.9813	<b>0.16</b>	0.14	-0.68
sv	Classic + MLN	Global	0.9613	0.0	<b>0.02</b>	0.0	0.9833	0.0	<b>0.01</b>	0.0	0.9578	0.0	<b>0.01</b>	0.0	0.9493	0.0	<b>0.02</b>	0.0
		Personalized	0.9613	0.0	<b>0.02</b>	0.0	0.9833	0.0	<b>0.01</b>	0.0	0.9578	0.0	<b>0.01</b>	0.0	0.9493	0.0	<b>0.02</b>	0.0
	Classic + MLN - Att	Global	0.9613	0.0	0.0	-16.42	0.9833	0.0	0.0	-12.46	0.9578	0.0	0.0	-24.19	0.9493	0.0	0.0	-22.96
		Personalized	0.9613	0.0	0.0	-16.42	0.9833	0.0	0.0	-12.46	0.9578	0.0	0.0	-24.19	0.9493	0.0	0.0	-22.96
Total				8	<b>9</b>	0		4	<b>11</b>	0		5	<b>10</b>	0		8	<b>9</b>	0

Tableau 7 – Gain d’impacte des modélisations pour le jeu de données AFB

CREDIT_RISK_DATASET			Accuracy				Precision				Recall				F1-score			
			classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^
xgb	Classic + MLN	Global	0.9263	<b>0.08</b>	0.02	-0.14	0.938	<b>0.12</b>	-0.13	-0.58	0.847	<b>0.17</b>	<b>0.17</b>	-0.19	0.8804	<b>0.12</b>	0.07	-0.24
		Personalized	0.9263	0.02	<b>0.08</b>	-0.25	0.938	0.0	<b>0.27</b>	-0.87	0.847	0.17	<b>0.18</b>	-0.35	0.8804	0.07	<b>0.14</b>	-0.41
	Classic + MLN - Att	Global	0.9263	<b>0.12</b>	-0.21	-0.36	0.938	0.0	-0.86	-1.1	0.847	<b>0.48</b>	-0.02	-0.58	0.8804	<b>0.27</b>	-0.27	-0.62
		Personalized	0.9263	<b>0.17</b>	0.02	-0.36	0.938	0.0	<b>0.18</b>	-1.27	0.847	<b>0.45</b>	-0.01	-0.52	0.8804	<b>0.33</b>	0.02	-0.6
dtc	Classic + MLN	Global	0.8825	-0.03	<b>0.1</b>	-0.29	0.7289	-0.25	<b>0.34</b>	-0.93	0.827	0.1	<b>0.13</b>	-0.39	0.827	-0.01	<b>0.15</b>	-0.45
		Personalized	0.8825	<b>-0.06</b>	-0.09	-0.65	0.7289	0.11	<b>0.23</b>	-2.39	0.827	-0.13	<b>0.25</b>	-0.57	0.827	-0.11	<b>-0.07</b>	-0.86
	Classic + MLN - Att	Global	0.8825	<b>-0.22</b>	-0.32	-0.74	0.7289	<b>-0.88</b>	-1.41	-2.96	0.827	-0.12	<b>-0.1</b>	-0.44	0.827	<b>-0.25</b>	-0.35	-0.89
		Personalized	0.8825	<b>0.24</b>	-0.09	-0.35	0.7289	<b>0.82</b>	-0.74	-1.92	0.827	0.28	<b>0.37</b>	0.16	0.827	<b>0.35</b>	-0.02	-0.3
lrc	Classic + MLN	Global	0.8489	<b>0.01</b>	0.0	-0.05	0.7311	<b>0.07</b>	0.0	-0.22	0.7152	<b>0.04</b>	0.0	-0.03	0.7434	<b>0.04</b>	0.0	-0.05
		Personalized	0.8489	-0.02	<b>0.04</b>	-0.08	0.7311	-0.04	<b>0.08</b>	-0.49	0.7152	-0.06	<b>0.1</b>	-0.03	0.7434	-0.05	<b>0.09</b>	-0.07
	Classic + MLN - Att	Global	0.8489	<b>-0.02</b>	-0.13	-1.85	0.7311	0.37	<b>0.71</b>	-3.1	0.7152	<b>-0.06</b>	-0.85	-5.8	0.7434	<b>-0.05</b>	-0.66	-5.54
		Personalized	0.8489	0.0	-0.49	-1.92	0.7311	0.52	<b>0.7</b>	-3.45	0.7152	0.0	-1.58	-5.93	0.7434	0.0	-1.47	-5.68
rfc	Classic + MLN	Global	0.9162	0.21	<b>0.35</b>	0.22	0.9296	0.0	<b>0.33</b>	-0.34	0.825	0.55	<b>0.84</b>	0.7	0.8616	0.42	<b>0.7</b>	0.49
		Personalized	0.9162	0.13	<b>0.22</b>	0.13	0.9296	<b>0.06</b>	-0.03	-0.3	0.825	0.36	<b>0.64</b>	0.46	0.8616	0.27	<b>0.48</b>	0.31
	Classic + MLN - Att	Global	0.9162	<b>0.19</b>	0.13	-0.01	0.9296	<b>0.09</b>	-0.12	-1.25	0.825	0.47	<b>0.55</b>	0.39	0.8616	<b>0.37</b>	0.31	0.1
		Personalized	0.9162	0.25	0.09	<b>0.35</b>	0.9296	0.57	-0.14	<b>0.98</b>	0.825	<b>0.64</b>	0.29	0.62	0.8616	0.48	0.19	<b>0.63</b>
sv	Classic + MLN	Global	0.8502	0.0	0.0	<b>0.02</b>	0.7783	0.0	-0.15	-0.14	0.6992	0.0	0.07	<b>0.11</b>	0.7329	0.0	0.05	<b>0.11</b>
		Personalized	0.8502	0.0	0.0	-0.07	0.7783	0.0	<b>0.1</b>	-0.59	0.6992	0.0	<b>0.04</b>	-0.03	0.7329	0.0	<b>0.01</b>	-0.07
	Classic + MLN - Att	Global	0.8502	0.0	-0.55	-2.09	0.7783	0.0	-0.48	-4.09	0.6992	0.0	-1.39	-6.21	0.7329	0.0	-1.45	-6.44
		Personalized	0.8502	0.0	-0.72	-2.11	0.7783	0.0	<b>0.69</b>	-4.33	0.6992	0.0	-2.43	-6.18	0.7329	0.0	-2.36	-6.43
Total				<b>9</b>	5	2		6	<b>10</b>	1		6	<b>11</b>	1		<b>8</b>	7	2

Tableau 8 – Gain d’impacte des modélisations pour le jeu de données CREDIT RISK DATASET

GERMAN		Accuracy				Precision				Recall				F1-score				
		classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	
xgb	Classic + MLN	Global	0.745	4.03	<b>7.38</b>	1.34	0.7021	6.82	<b>13.94</b>	4.44	0.6819	5.79	<b>9.59</b>	1.13	0.6919	6.0	<b>10.12</b>	1.32
		Personalized	0.745	3.36	<b>4.7</b>	0.0	0.7021	8.37	<b>15.3</b>	0.0	0.6819	4.27	<b>5.88</b>	0.0	0.6919	4.57	<b>6.3</b>	0.0
	Classic + MLN - Att	Global	0.745	<b>6.04</b>	5.37	-6.71	0.7021	<b>14.56</b>	13.3	-16.22	0.6819	7.01	<b>7.41</b>	-7.1	0.6919	7.67	<b>7.72</b>	-7.7
		Personalized	0.745	6.04	<b>6.71</b>	-11.41	0.7021	14.56	<b>15.3</b>	-25.55	0.6819	7.01	<b>8.46</b>	-14.43	0.6919	7.67	<b>8.73</b>	-15.75
dtc	Classic + MLN	Global	0.62	<b>10.48</b>	<b>10.48</b>	8.06	0.45	<b>24.29</b>	<b>24.29</b>	17.84	0.5659	12.72	<b>14.47</b>	11.47	0.567	12.8	<b>13.99</b>	11.53
		Personalized	0.62	8.06	<b>10.48</b>	4.03	0.45	18.04	<b>23.02</b>	9.36	0.5659	10.89	<b>14.67</b>	5.74	0.567	11.06	<b>14.69</b>	5.84
	Classic + MLN - Att	Global	0.62	<b>10.48</b>	<b>10.48</b>	-1.61	0.45	23.87	<b>24.29</b>	-1.96	0.5659	12.92	<b>13.5</b>	0.39	0.567	13.3	<b>13.79</b>	0.26
		Personalized	0.62	8.06	<b>11.29</b>	2.42	0.45	17.84	<b>24.18</b>	5.27	0.5659	11.47	<b>15.92</b>	2.05	0.567	11.53	<b>15.87</b>	2.1
lrc	Classic + MLN	Global	0.745	0.0	<b>0.67</b>	0.0	0.7111	0.0	<b>2.28</b>	0.0	0.6786	0.0	<b>0.56</b>	0.0	0.6888	0.0	<b>0.65</b>	0.0
		Personalized	0.745	0.0	0.0	0.0	0.7111	0.0	0.0	0.0	0.6786	0.0	0.0	0.0	0.6888	0.0	0.0	0.0
	Classic + MLN - Att	Global	0.745	0.67	<b>2.01</b>	-14.09	0.7111	0.89	<b>8.99</b>	-43.75	0.6786	1.05	<b>2.18</b>	-25.51	0.6888	1.12	<b>2.44</b>	-36.48
		Personalized	0.745	2.01	<b>4.03</b>	-14.09	0.7111	<b>18.66</b>	11.19	-41.4	0.6786	1.62	<b>4.36</b>	-25.02	0.6888	1.77	<b>4.89</b>	-34.99
rfc	Classic + MLN	Global	0.73	<b>4.79</b>	<b>4.79</b>	0.0	0.75	<b>13.13</b>	10.48	-1.96	0.6407	6.26	<b>6.77</b>	0.52	0.6447	7.45	<b>8.03</b>	0.7
		Personalized	0.73	2.05	<b>2.74</b>	-1.37	0.75	3.23	<b>7.53</b>	-5.88	0.6407	4.37	<b>4.96</b>	-1.2	0.6447	5.23	<b>5.93</b>	-1.32
	Classic + MLN - Att	Global	0.73	3.42	<b>6.85</b>	-1.37	0.75	4.51	<b>12.28</b>	-8.77	0.6407	5.57	<b>10.11</b>	-0.17	0.6447	6.62	<b>11.83</b>	0.03
		Personalized	0.73	<b>4.11</b>	2.74	-6.16	0.75	<b>12.51</b>	5.05	-17.24	0.6407	5.14	<b>5.48</b>	-8.49	0.6447	6.13	<b>6.48</b>	-10.47
sv	Classic + MLN	Global	0.755	0.66	<b>1.32</b>	0.0	0.7561	2.5	<b>5.13</b>	0.0	0.683	0.56	<b>1.61</b>	0.0	0.6945	0.66	<b>1.8</b>	0.0
		Personalized	0.755	<b>0.66</b>	<b>0.66</b>	0.0	0.7561	<b>2.5</b>	<b>2.5</b>	0.0	0.683	<b>0.56</b>	<b>0.56</b>	0.0	0.6945	<b>0.66</b>	<b>0.66</b>	0.0
	Classic + MLN - Att	Global	0.755	0.66	<b>1.99</b>	-13.91	0.7561	2.5	<b>6.55</b>	-100.0	0.683	0.56	<b>3.13</b>	-26.79	0.6945	0.66	<b>3.35</b>	-43.28
		Personalized	0.755	0.66	<b>3.31</b>	-13.91	0.7561	3.32	<b>7.23</b>	-100.0	0.683	0.97	<b>4.26</b>	-26.79	0.6945	0.94	<b>4.71</b>	-43.28
Total				6	<b>17</b>	0		6	<b>15</b>	0		1	<b>19</b>	0		1	<b>19</b>	0

Tableau 9 – Gain d’impacte des modélisations pour le jeu de données GERMAN

JAPAN		Accuracy				Precision				Recall				F1-score				
		classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	classic	MLN 1^	MLN 2^	MLN All^	
xgb	Classic + MLN	Global	0.8551	1.7	<b>3.39</b>	<b>3.39</b>	0.913	0.27	<b>1.82</b>	-1.85	0.8599	1.51	<b>3.22</b>	2.63	0.8546	1.67	<b>3.37</b>	3.26
		Personalized	0.8551	1.7	<b>5.09</b>	2.54	0.913	0.27	<b>2.03</b>	-2.0	0.8599	1.51	<b>4.73</b>	1.87	0.8546	1.67	<b>5.05</b>	2.42
	Classic + MLN - Att	Global	0.8551	3.39	<b>5.09</b>	-6.78	0.913	0.53	<b>2.03</b>	-9.45	0.8599	3.02	<b>4.73</b>	-7.43	0.8546	3.35	<b>5.05</b>	-6.97
		Personalized	0.8551	2.54	<b>3.39</b>	1.7	0.913	0.41	<b>1.82</b>	-3.27	0.8599	2.27	<b>3.22</b>	0.92	0.8546	2.5	<b>3.37</b>	1.54
dtc	Classic + MLN	Global	0.8116	<b>3.57</b>	<b>3.57</b>	1.79	0.84	<b>4.37</b>	<b>4.37</b>	1.58	0.8107	<b>3.84</b>	<b>3.84</b>	1.81	0.8096	<b>3.68</b>	<b>3.68</b>	1.82
		Personalized	0.8116	2.67	<b>4.46</b>	-7.15	0.84	2.96	<b>5.44</b>	-5.42	0.8107	2.82	<b>4.22</b>	-7.03	0.8096	2.74	<b>4.43</b>	-7.15
	Classic + MLN - Att	Global	0.8116	2.67	<b>6.25</b>	-9.82	0.84	2.74	<b>6.18</b>	-9.95	0.8107	2.41	<b>6.45</b>	-10.28	0.8096	2.62	<b>6.36</b>	-10.1
		Personalized	0.8116	4.46	<b>5.36</b>	-3.57	0.84	4.37	<b>5.44</b>	-3.18	0.8107	4.43	<b>5.02</b>	-3.63	0.8096	4.5	<b>5.31</b>	-3.62
lrc	Classic + MLN	Global	0.8551	0.0	0.0	0.0	0.913	0.0	0.0	0.0	0.8599	0.0	0.0	0.0	0.8546	0.0	0.0	0.0
		Personalized	0.8551	0.0	0.0	0.0	0.913	0.0	0.0	0.0	0.8599	0.0	0.0	0.0	0.8546	0.0	0.0	0.0
	Classic + MLN - Att	Global	0.8551	0.0	0.0	-6.78	0.913	0.0	0.0	-12.1	0.8599	0.0	0.0	-8.02	0.8546	0.0	0.0	-7.24
		Personalized	0.8551	0.0	0.0	-8.48	0.913	0.0	<b>1.36</b>	-13.46	0.8599	0.0	<b>0.2</b>	-9.73	0.8546	0.0	<b>0.02</b>	-8.97
rfc	Classic + MLN	Global	0.8768	0.0	<b>0.83</b>	-0.82	0.9167	0.0	<b>1.3</b>	-1.37	0.8794	0.0	<b>0.55</b>	-0.93	0.876	0.0	<b>0.79</b>	-0.84
		Personalized	0.8768	<b>0.83</b>	<b>0.83</b>	0.9167	<b>1.41</b>	0.12	0.12	0.8794	<b>0.93</b>	0.74	0.74	0.876	<b>0.84</b>	0.82	0.82	
	Classic + MLN - Att	Global	0.8768	0.0	<b>0.83</b>	-7.44	0.9167	0.0	<b>0.12</b>	-11.12	0.8794	0.0	<b>0.74</b>	-8.39	0.876	0.0	<b>0.82</b>	-7.81
		Personalized	0.8768	0.0	<b>0.83</b>	-1.65	0.9167	0.0	<b>0.12</b>	-3.83	0.8794	0.0	<b>0.74</b>	-2.06	0.876	0.0	<b>0.82</b>	-1.75
sv	Classic + MLN	Global	0.8478	0.0	0.0	0.0	0.9242	0.0	0.0	0.0	0.8551	0.0	0.0	0.0	0.8476	0.0	0.0	0.0
		Personalized	0.8478	0.0	0.0	0.0	0.9242	0.0	0.0	0.0	0.8551	0.0	0.0	0.0	0.8476	0.0	0.0	0.0
	Classic + MLN - Att	Global	0.8478	<b>0.86</b>	0.0	-6.83	0.9242	0.0	0.0	-14.99	0.8551	<b>0.36</b>	0.0	-8.67	0.8476	<b>0.79</b>	0.0	-7.52
		Personalized	0.8478	0.0	<b>0.86</b>	-6.83	0.9242	0.0	<b>1.55</b>	-15.7	0.8551	0.0	<b>0.96</b>	-8.86	0.8476	0.0	<b>0.87</b>	-7.66
Total				3	<b>13</b>	2		2	<b>13</b>	0		3	<b>13</b>	0		3	<b>13</b>	0

Tableau 10 – Gain d’impacte des modélisations pour le jeu de données JAPAN

Les figure ci-contre nous montre que dans certains cas, la modélisation à 1 couche peut permettre d’extraire des motifs du graphe qui améliores les métriques de classification. Cependant il apparait clairement que le fait de modéliser une relation des emprunts sur la base de toutes les attributs catégoriels possible n’est pas vraiment une solution car n’améliore pas vraiment les modèles classiques. Il serait donc judicieux de pouvoir identifier quel sont les attributs catégoriels qui garantissent d’apporter de meilleurs résultats

## 4.4 Choix des attributs à considérer

Le choix optimal des attributs devant servir à modéliser nos données de prêts peut vraiment avoir de l'impact sur les modèle prédictif résultant et même la qualité des motifs extraits des graphes construits. Pour acter ce besoin, nous allons sur l'intuition de la disponibilité de correlation entre l'impact de la modélisation d'un attribut en MLN 1 et la qualité des couplage d'ordre  $k_{>=2}$ . Dans un sens, nous avons proposé un clustering des impacts de modélisation à une couche pour voire observer la co-occurrence des comportements des couplages à la couche  $k \leftarrow 2$  dans le cas de nos expérimentations.

accuracy			(Good MLN 1, Good MLN 1)			(Good MLN 1, Bad MLN 1)			(Bad MLN 1, Bad MLN 1)		
			Classic >	Classic =	Classic <	Classic >	Classic =	Classic <	Classic >	Classic =	Classic <
Global	Classic + MLN	21	18	107	28	21	59	124	164	78	
	Classic + MLN - Att	20	12	95	73	28	85	173	59	75	
Personalized	Classic + MLN	15	14	119	31	31	77	61	222	50	
	Classic + MLN - Att	35	16	106	81	18	76	160	59	69	
precision			(Good MLN 1, Good MLN 1)			(Good MLN 1, Bad MLN 1)			(Bad MLN 1, Bad MLN 1)		
			Classic >	Classic =	Classic <	Classic >	Classic =	Classic <	Classic >	Classic =	Classic <
Global	Classic + MLN	22	6	104	31	8	49	174	155	71	
	Classic + MLN - Att	25	5	94	79	6	74	210	57	70	
Personalized	Classic + MLN	16	5	112	44	24	67	103	207	42	
	Classic + MLN - Att	39	5	99	107	8	76	165	49	72	
recall			(Good MLN 1, Good MLN 1)			(Good MLN 1, Bad MLN 1)			(Bad MLN 1, Bad MLN 1)		
			Classic >	Classic =	Classic <	Classic >	Classic =	Classic <	Classic >	Classic =	Classic <
Global	Classic + MLN	32	6	113	35	10	73	119	150	82	
	Classic + MLN - Att	30	5	121	89	13	95	165	48	54	
Personalized	Classic + MLN	27	7	137	38	21	79	61	204	46	
	Classic + MLN - Att	29	4	126	90	6	83	161	50	71	
f1-score			(Good MLN 1, Good MLN 1)			(Good MLN 1, Bad MLN 1)			(Bad MLN 1, Bad MLN 1)		
			Classic >	Classic =	Classic <	Classic >	Classic =	Classic <	Classic >	Classic =	Classic <
Global	Classic + MLN	37	7	114	36	9	75	114	150	78	
	Classic + MLN - Att	29	5	122	87	13	100	163	48	53	
Personalized	Classic + MLN	28	7	139	37	21	77	61	204	46	
	Classic + MLN - Att	31	4	127	85	6	86	161	50	70	

Tableau 11 – Correlation entre l'impact d'attributs en MLN 1 et celui de des couples en MLN 2 pour tous les jeux de données

Nous remarquons suivant la figure 15, nous remarquons que les couples formés d'attributs ayant un bon impact en MLN 1 (Good MLN 1, Good MLN 1) ont tendances à améliorer les modèles classiques en MLN 2.

## 4.5 Les meilleures performances

Des résultats des sections 4.2, 4.3, nous pouvons extraire les tendances suivantes :

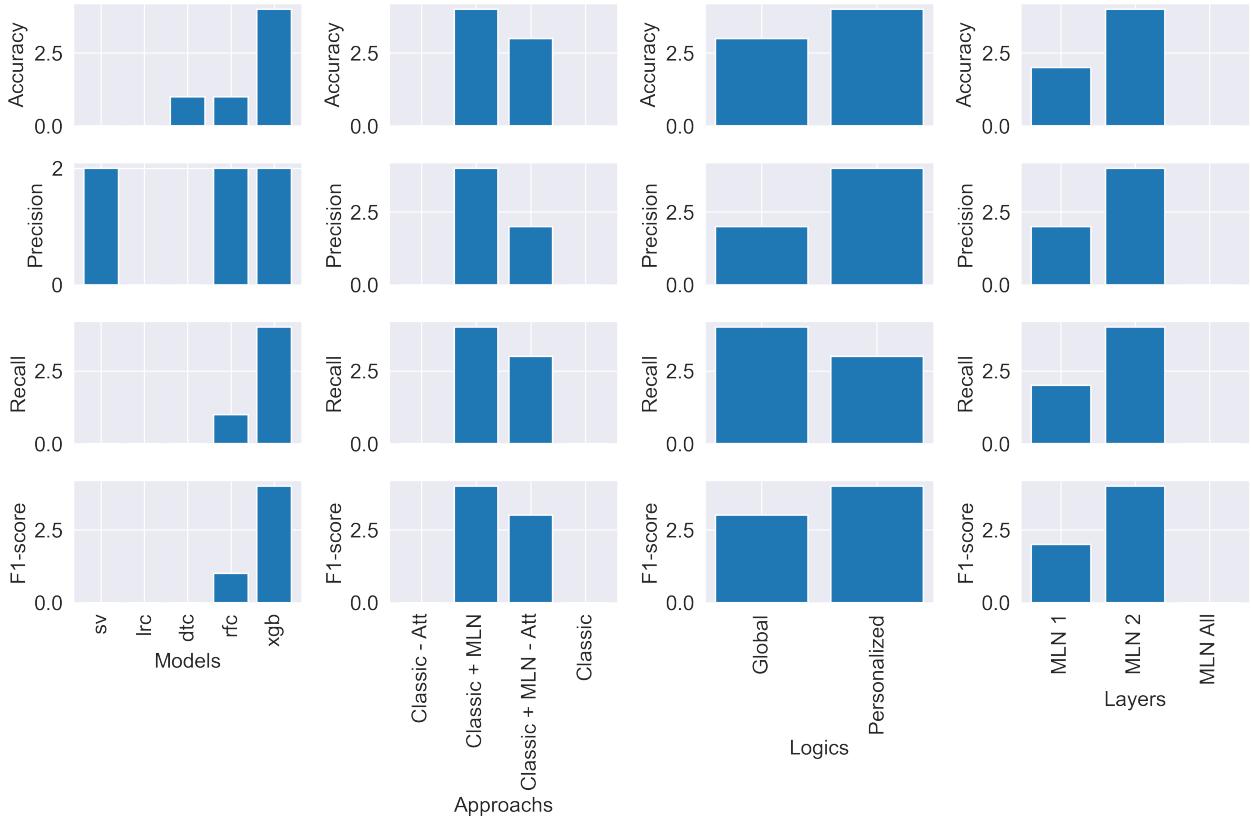


Figure 16 – Importance des modèles, logiques, approches, couches dans l'évaluation du risque de crédit.

Premièrement le modèle XGB est le modèle qui généralise au mieux l'ensemble des données des différent jeux de données. Ensuite **Classic + MLN** permet d'optimiser les métriques sur les différents jeux de données. Puis nous remarquons que **la personnalisation à un seul emprunteur** surpassé la personnalisation global sur nos jeux de données. Enfin la modélisation d'un **MLN 2** permet de mieux capturer des motifs dans les graphes issus des données de prêts des différents jeux de données.

## 5 Conclusion et perspectives

Au terme de cette rédaction, il était question pour nous de présenter les différentes expérimentation réalisées au cours de ses derniers mois et leur résultats. Il en ressort que l'idée de personnalisation à un seul emprunteur apparaît comme un intuition valide mais l'apport n'est pas très éloigné de la version global proposé par les auteurs de l'article de base mais s'adapte mieux sur tout jeu de données. La modélisation à une couche nous aura permis de comprendre que si nous voulons construire des  $MLN_{k \geq 2}$ , il serait judicieux de prendre ceux là donc la version **MLN 1** améliore les modèles classiques.

Lors de ces expérimentation, Nous avons remarqué la non importance de plusieurs attributs des jeux de données dans les modèles, l'absence d'information de classe dans le graphe multicouche biparti construit, la minorité d'attributs extraits des graphes. Pour cela nous pensons au perspectives suivantes :

1. définir un processus de présélection des attributs avant entraînement
2. intégrer les classes dans la modélisation du graphe multicouche
3. exploitation différente du graphe multicouche sans avoir recourt au PageRank.

## Références

- [1] M. Oskarsdottir and C. Bravo, “Multilayer network analysis for improved credit risk prediction,” *Omega*, vol. 105, p. 102520, 2021.
- [2] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of modern physics*, vol. 87, no. 3, pp. 925—979, 2015.
- [3] S. Mario, “Représentations graphiques des portefeuilles de crédit et leur analyse,” *Revue européenne des sciences économiques et de gestion*, no. 1, pp. 23–28, 2021.
- [4] X. Xu, C. Zhou, and Z. Wang, “Credit scoring algorithm based on link analysis ranking with support vector machine,” *Expert systems with Applications*, vol. 36, no. 2, pp. 2625–2632, 2009.
- [5] S. Shi, R. Tse, W. Luo, S. D’Addona, and G. Pau, “Machine learning-driven credit risk : a systemic review,” *Neural Computing and Applications*, vol. 34, no. 17, pp. 14 327–14 339, 2022.
- [6] M. Yıldırım, F. Y. Okay, and S. Özdemir, “Big data analytics for default prediction using graph theory,” *Expert Systems with Applications*, vol. 176, p. 114840, 2021.
- [7] C. Wang, F. Yu, Z. Zhang, and J. Zhang, “Multiview graph learning for small-and medium-sized enterprises’ credit risk assessment in supply chain finance,” *Complexity*, vol. 2021, pp. 1–13, 2021.
- [8] S. U. Rehman, A. U. Khan, and S. Fong, “Graph mining : A survey of graph mining techniques,” in *Seventh International Conference on Digital Information Management (ICDIM 2012)*. IEEE, 2012, pp. 88–92.
- [9] D. A. Vega-Oliveros, P. S. Gomes, E. E. Milios, and L. Berton, “A multi-centrality index for graph-based keyword extraction,” *Information Processing & Management*, vol. 56, no. 6, pp. 4–5, 2019.
- [10] L. F. P. Tous, “Le crédit,” Jan. 2023. [Online]. Available : <https://www.lafinancepourtous.com/decryptages/marches-financiers/acteurs-de-la-finance/banque/la-banque-a-quoi-ca-sert/le-credit/>
- [11] ——, “La crise des subprimes (2007-2008),” Jan. 2024. [Online]. Available : <https://www.lafinancepourtous.com/juniors/lyceens/les-crises/la-crise-des-subprimes-2007-2008/>