

# PRÉDICTION DU RISQUE DE CRÉDIT EN INTÉGRANT LES DESCRIPTEURS EXTRAITS DE LA MODÉLISATION EN GRAPHE DES DONNÉES BANCAIRES

RÉDIGÉ PAR

VICTOR NICO DJIEMBOU TIENTCHEU  
ARMEL JACQUES NZEKON NZEKO'O

## RÉSUMÉ

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. La plupart des travaux sur la prédition du risque de crédit considèrent uniquement les attributs de description des prêts, 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.

L'objectif de ce travail est d'explorer les voies d'utilisation de la modélisation des données bancaires par des graphes afin d'améliorer la qualité des modèles classiques de prédition du risque de crédit. Nous nous attardons sur le cas des graphes multicouches dont l'usage nécessite d'abord la modélisation du graphe en considérant des attributs catégoriels comme couches de ces graphes, ensuite l'application du PageRank pour extraire des caractéristiques qui sont ajoutés en entrée du processus d'apprentissage des modèles de prédition.

La compréhension des travaux existant nous permet de les répliquer sur un jeu de données d'Afriland First Bank et sur quatre autres jeux de données standards dédiés à la prédition du risque de crédit, ce qui a permis de confirmer l'importance des descripteurs extraits des graphes multicouches. Par ailleurs, nous avons abordé des pistes de solutions aux limites des travaux existants notamment en proposant de Personnaliser le PageRank pour chaque prêt au lieu d'avoir une seule exécution PageRank pour tous les prêts. La pertinence de cette première contribution s'illustre par le fait dans 81% des cas, l'attribut le plus important parmi ceux extraits du graphe multicouches est issu de la Personnalisation du PageRank pour chaque prêt.

De plus, nous avons évalué l'impact du nombre d'attributs considérés comme couches dans les graphes multicouches, et il en ressort que ni les graphes multicouches associé un seul attribut, ni le graphe multicouches associé à tous les attributs catégoriels ne permet d'avoir généralement les meilleures prédition du risque de crédit, mais c'est le graphe multicouches à deux couches qui le permet en général. Le constat précédent a confirmé la nécessité de procéder à une sélection des attributs à considérer dans le graphe multicouche, contrairement aux travaux existants qui n'en propose aucune. Ainsi, nous avons proposé un protocole pour le choix des attributs de construction des graphes multicouches à k couches, dont le test sur les graphes multicouches à deux couches a été un succès à plus de 75%.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Descripteurs extraits des graphes</b>	<b>2</b>
2.1	Généralité et définition . . . . .	2
2.2	Travaux existants sur l'extraction des descripteurs des graphes . . . . .	3
<b>3</b>	<b>Descripteurs extraits des graphes multicouches</b>	<b>3</b>
3.1	Processus de construction du graphe . . . . .	4
3.2	Extraction des nouveaux descripteurs du graphe multicouches . . . . .	5
3.3	Prédiction du risque de crédit avec les nouveaux descripteurs . . . . .	6
3.4	Limites et perspectives au travail sur le graphe multicouches . . . . .	6
<b>4</b>	<b>Expérimentations et résultats</b>	<b>7</b>
4.1	Jeux de Données . . . . .	7
4.2	Application des descripteurs du graphe multicouches sur un jeu de données d'AfriLand First Bank . . . . .	7
4.2.1	Mise en œuvre : intégration des attributs extraits des graphes multicouches dans le processus de prédiction du risque de crédit . . . . .	8
4.2.2	Évaluation de la contribution des attributs pour les modèles de prédiction	10
4.3	Descripteurs génériques extraits des graphes multicouches MLNk . . . . .	16
4.4	Application du PageRank personnalisé pour chaque prêt . . . . .	17
4.5	Choix des attributs à considérer dans le graphe multicouches . . . . .	25
4.5.1	Graphes multicouches avec des nombres différents de couches . . . . .	26
4.5.2	Sélection des attributs à considérer dans un graphe multicouches . . . . .	28
4.6	Configurations qui permettent d'avoir les meilleures performances . . . . .	30
<b>5</b>	<b>Conclusion et perspectives</b>	<b>31</b>
5.1	Conclusion . . . . .	31
5.2	Perspectives . . . . .	32

# 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 [1, 2, 3].

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 nœuds 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 [4].

Le reste du document est structuré comme suit : nous débutons par la section 2 pour la présentation des généralités sur les graphes et les descripteurs extraits des graphes. Ensuite, nous présentons dans la section 3 la modélisation par des graphes multicouches et les descripteurs extraits de ces graphes. Enfin, nous présentons quelques expérimentations réalisées et les résultats obtenus à la section 4 et nous terminons par une conclusion sur le travail fournis 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 nœuds  $V$  et celle d'arcs  $E$ . La notion de nœud est associée aux entités qui sont en relation et celle d'arc est associée à la nature de la relation d'une entité (nœud) avec une autre.

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

- **Descripteurs d'un nœud :** on peut citer les mesures de centralité qui estiment à quel point le nœud est incontournable dans la navigation dans le graphe. C'est le cas par exemple du PageRank[5] où initialement on attribue le même poids à chaque nœud, puis chaque nœud 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 nœuds ne changent plus, ou alors jusqu'à ce qu'un nombre maximum d'étapes de diffusion soit atteint. Les nœuds aux poids les plus grands, sont les plus importants.
- **Descripteurs d'une relation entre deux nœuds :** on peut parler des mesures qui décrivent la relation entre deux nœuds à partir du nombre et de la longueur des plus courts chemins entre ces nœuds.

- **Descripteurs de la topologie du graphe :** il est possible d'extraire des communautés dans un graphe (sous-ensemble de nœuds 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.

## 2.2 Travaux existants sur l'extraction des descripteurs des graphes

Dans la littérature sur la prédiction du risque de crédit, quelques travaux procèdent à des modélisations des données de prêts par des graphes. A l'issu de la construction du graphe, des opérations y sont appliquées pour extraire de nouveaux descripteurs qui peuvent être pertinents.

Mario et al. [6] capturent les relations financières entre entités intervenantes (emprunteurs, instituts financiers) dans le contexte d'une inter-coopération des institutions financières et des emprunteurs pour modéliser un graphe orienté représentant une micro-structure du réseau où chaque noeud représente un emprunteur ou une institution financière, et les arêtes, le lien financier existant entre les noeuds. Les auteurs vont extraire des mesures telles que le degré d'un noeud, les longueurs des plus courts chemins, le sous graphe maximal de distance minimale afin d'améliorer la prédiction du risque de crédit avec des modèles comme SVM -*Support Vector Machine (SVM)* et la Régression Logistique -*Logistic Regression (LR)*.

Xu et al. [7] construisent un graphe biparti pour représenter les informations d'historique des prêts dans les institutions bancaires. Ici chaque noeud matérialise soit un emprunteur soit une modalité des attributs, et une arrête entre un noeud emprunteur et un noeud modalité signifie que l'emprunteur est décrit par cette modalité. Pour prendre en compte les relations complexe entre les emprunteurs, ils vont utiliser trois algorithmes d'analyse de liens basés sur Hub Authority Ranking Applicants Algorithm (HARA), Hub-Avg ranking applicants Algorithm (HubAvgRA), Authority-Threshold Algorithm (ATkRA) pour extraire des descripteurs qui sont positionnés en entrée de l'apprentissage du modèle SVM.

Giudici et al. [8] procèdent à la construction des graphes complets dans lesquels un noeud est un prêt ou un emprunteur, et les liens ou arêtes entre les noeuds sont pondérés par des valeurs qui correspondent à la similarité entre les deux noeuds associés. Une fois le graphe complet construit, plusieurs algorithmes d'extraction des caractéristiques sur des graphes sont appliqués (mesures de centralité des noeuds), et les résultats obtenus sont positionnés en entrée du processus d'apprentissage des modèles de prédiction comme nouveaux descripteurs.

Les graphes bipartis et les graphes complets ne sont pas les seuls types de graphes qui peuvent être exploités pour la modélisation des données bancaires. Ainsi, explorer d'autres modélisation en graphe dans le but d'extraire des attributs pertinents reste une voie intéressante pour les futurs travaux sur les descripteurs extraits des graphes. Une des pistes envisagée récemment est celle de la modélisation des données bancaires par des graphes multicouches afin de mieux capturer les relations entre emprunteurs, et entre emprunteur et leurs caractéristiques.

## 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 [4], est le point d'entrée pour l'usage des descripteurs issus de la modélisation par un graphe multicouches, avec pour objectif d'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 nœuds 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 chacune des dimensions du graphe multicouches. Une dimension peut par exemple être la localisation géographique ou encore le type d'activité exercé comme c'est le cas dans les travaux de Óskarsdóttir et al. [4].

L'une des hypothèses sur laquelle les auteurs s'appuient, est celle selon laquelle, les emprunteurs qui ont un grand nombre de caractéristiques en commun (dans les différentes dimensions) ont une grande probabilité d'avoir des prêts de la même classe (PAYE ou IMPAYE).

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 proposent de construire un graphe multicouches, et pour ce faire, on fixe les dimensions considérées et les attributs associés à chacune de ces dimensions. Dans le cas des travaux de Óskarsdóttir et al. [4] il est question de prêts agricoles et seules deux dimensions sont choisies : 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.

La construction du graphe multicouches obéit aux règles suivantes :

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

Dans l'article, les auteurs considèrent deux dimensions pour décrire les emprunteurs dans le graphe multicouches, à savoir la localité et les produits vendus par ces derniers. Considérons un cas où nous avons 4 emprunteurs (des fermiers), 2 localités et 3 produits agricoles vendus par les fermiers. Dans ce cas, nous avons 3 couches (Emprunteur, Localité et Produit), et nous avons 9 noeuds (4 noeuds emprunteurs + 2 noeuds localités + 3 noeuds produits), et donc la matrice carrée qui permet de représenter le graphe multicouches est de taille (9 x 3) x (9 x 3).

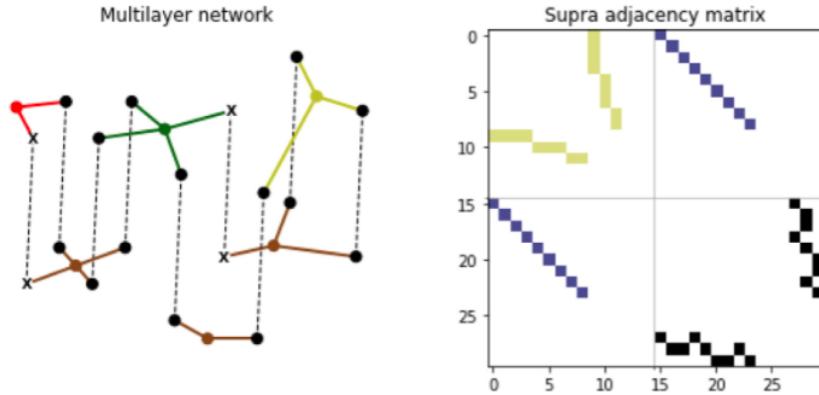


Figure 1 – A gauche on a un exemple de graphe multicouches à deux couches, et à droite on a la représentation de ce graphe sous forme de matrice. Ce graphe contient 9 emprunteurs (nœuds noirs), 3 nœuds de la dimension localité des emprunteurs (nœuds marron) et 3 nœuds de la dimension produits (vert, 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).

### 3.2 Extraction des nouveaux descripteurs du graphe multicouches

Lorsque le graphe multicouches est construit, les nouveaux descripteurs du prêt sont calculés suite à des applications du PageRank Personnalisé sur le graphe résultat. Ceci est possible car un graphe multicouches  $M$ , ayant  $N$  nœuds, et  $L$  couches, correspond à une représentation de dimension  $N \times N \times L \times L$ , ce qui peut être résumé en une matrice carrée  $(N \times L) \times (N \times L)$  sur laquelle on peut appliquer le PageRank. Ainsi, 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.

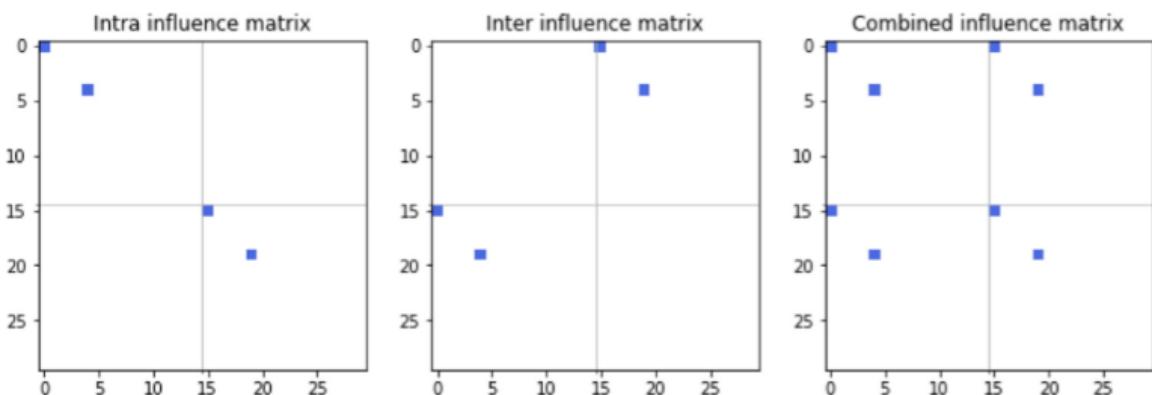


Figure 2 – Les trois scénarios considérés par les auteurs pour calculer les nouveaux descripteurs des prêts par l'application du PageRank Personnalisé sur le graphe multicouches.

Notons que d'autres descripteurs sont considérés dans le graphe multicouches :

- Nombre de nœuds emprunteurs qui vendent les mêmes produits que l'emprunteur cible
- Nombre de nœuds emprunteurs défaillants qui vendent les mêmes produits que l'emprunteur cible
- Nombre de nœuds emprunteurs de la même localité que l'emprunteur cible
- Nombre de nœuds 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 (Régression 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 Régression logistique et XGBoost.

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

Óskarsdóttir et al. [4] proposent une modélisation des données des prêts en graphe multicouches. 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 multicouches.
- Le choix des attributs catégoriels à considérer comme couche du graphe multicouches est fait de façon arbitraire. Il serait intéressant de proposer un protocole pour ce choix.
- Les applications du PageRank Personnalisé sur le graphe multicouches sont globales à tout le graphe, et donc tous les noeuds sont traités pareil quelque soit le prêt pour lequel on veut faire une prédiction. Et pourtant, il serait intéressant d'avoir une personnalisation du PageRank qui favorise les noeuds qui sont liés ou qui sont proches de l'emprunteur dans le graphe multicouches. Par contre, on peut avoir des applications du PageRank personnalisé pour chacun des prêts ou emprunteur.

**Extensions possibles :** Ses critiques permettent d'émettre les extensions possibles suivantes :

- Proposer un graphe multicouches avec autant de couches que d'attributs catégoriels.
- Intégrer les attributs quantitatifs comme des couches du graphe multicouches.
- Proposer un protocole qui permet de choisir efficacement les attributs catégoriels à considérer dans le graphe multicouches.
- Personnaliser l'exécution du PageRank pour chaque prêt au lieu d'une exécution pour tous les prêts à la fois.
- Intégrer la classe du prêt comme une couche du graphe multicouches.
- Travailler avec plusieurs jeux de données.

## 4 Expérimentations et résultats

### 4.1 Jeux de Données

Pour explorer davantage l'impact des descripteurs extraits des graphes multicouches, en plus du jeu de données AFB de Afriland First Bank, nous considérons 04 jeux de données classiques dédiés à la prédiction du risque de crédit bancaire : AER, GERMAN, JAPAN, CREDIT\_RISK. Ces jeux données sont accessibles depuis les liens suivants :

- <https://github.com/JLZml/Credit-Scoring-Data-Sets>
- <https://www.kaggle.com/datasets/laotse/credit-risk-dataset>

Les détails sur les jeux de données considérés sont présents dans le tableau 1, avec comme première colonne l'identifiant du jeu de données, suivi respectivement du nombre de prêts, le nombre d'attributs descriptifs de chaque prêt, le nombre d'attributs numériques, le nombre d'attributs catégoriels et enfin les nombres de Bon et Mauvais prêt pour la Banque.

	Nb exemples	Nb attributs (Att)	Att numériques	Att catégoriels	Nb positifs(+)	Nb négatifs (-)
AFB	28952	13	08	04	21769	7183
AER	1319	12	09	02	1023	296
CREDIT RISK	32581	11	07	03	25473	7108
GERMAN	1000	21	07	13	700	300
JAPAN	690	16	06	09	307	383

TABLE 1 – Description des jeux de données

### 4.2 Application des descripteurs du graphe multicouches sur un jeu de données d'Afriland First Bank

Pour expérimenter les concepts appris, il est nécessaire de choisir des attributs qualitatifs qui vont représenter les dimensions (couches) du graphe multicouches. 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 des ces différents graphes multicouches à deux couches proposés par les auteurs sont énumérés comme suit pour une paire d'attribut (Att1, Att2) :

- **MLN\_Att1\_degré** : le nombre d'emprunteurs avec la même valeur d'Att
- **MLN\_Att2\_degré** : le nombre d'emprunteurs avec la même valeur d'Att
- **MLN\_Att1\_et\_Att2\_degré** : le nombre d'emprunteurs qui ont des valeurs égales pour chacun des attributs 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

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

Nous avons considéré cinq modèles classiques de l'apprentissage automatique pour la prédition 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 prédition 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 MultiLayer Network (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 multicouches MLN, et on intègre les attributs extraits du graphe multicouches 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 2 ci-dessous présente l'ensemble des résultats obtenus pour les cinq différents modèles de prédition 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

TABLE 2 – Résultats avec les différents graphes multicouches. Dans ce tableau, AMLNi correspond à Att et donc aux attributs considérés pour construire le graphe 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 ferme ce classement. En s'intéressant aux meilleures performances obtenues avec les attributs extraits des graphes multicouches, on constate que le graphe multicouches MLN1 est le meilleur pour le modèles 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 multicouches qui lui correspond le mieux, on fait les remarques suivantes :

- **SVM** : il ne faut pas considérer les attributs issus du graphe multicouches
- **XGBoost** : les meilleures performances sont atteintes lorsqu'on considère à la fois les attributs extraits du graphe multicouches 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 multicouches.
- **Arbre de décision** : 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.
- **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.

#### 4.2.2 Évaluation 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 dans leur forme 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 multicouches.

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.

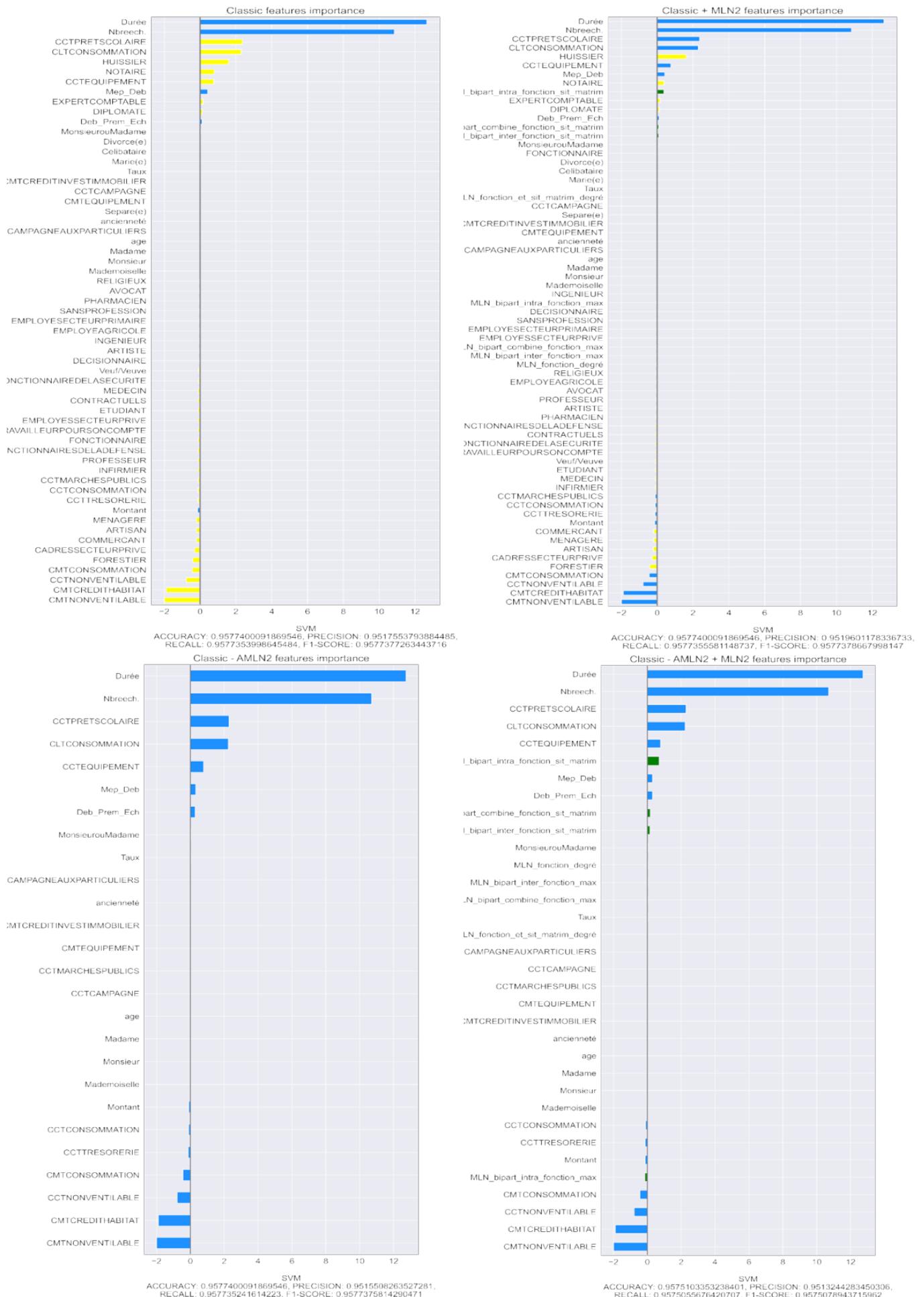


FIGURE 3 – SVM – meilleur graphe multicouches MLN2

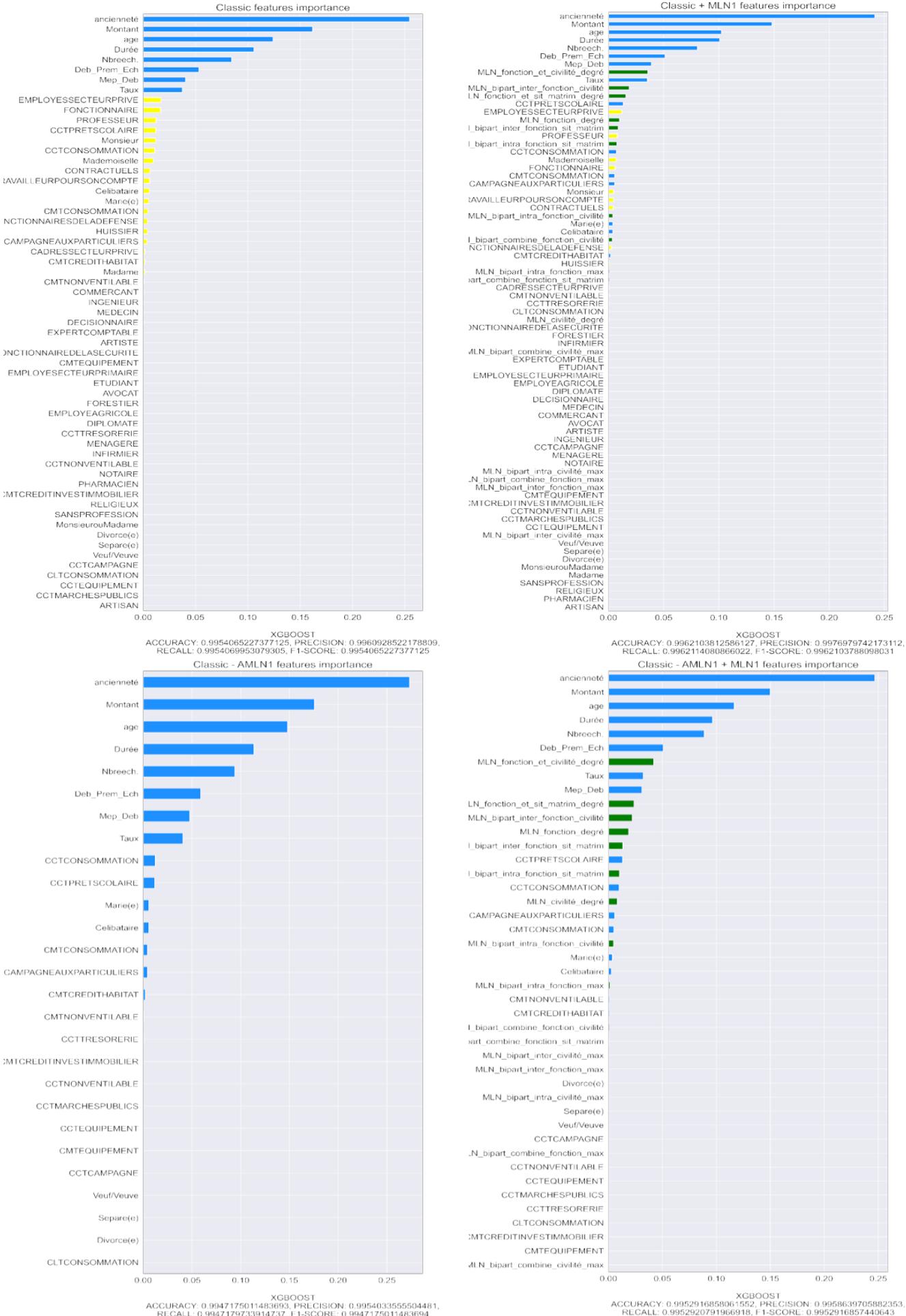


FIGURE 4 – XGBoost – meilleur graphe multicouches MLN1

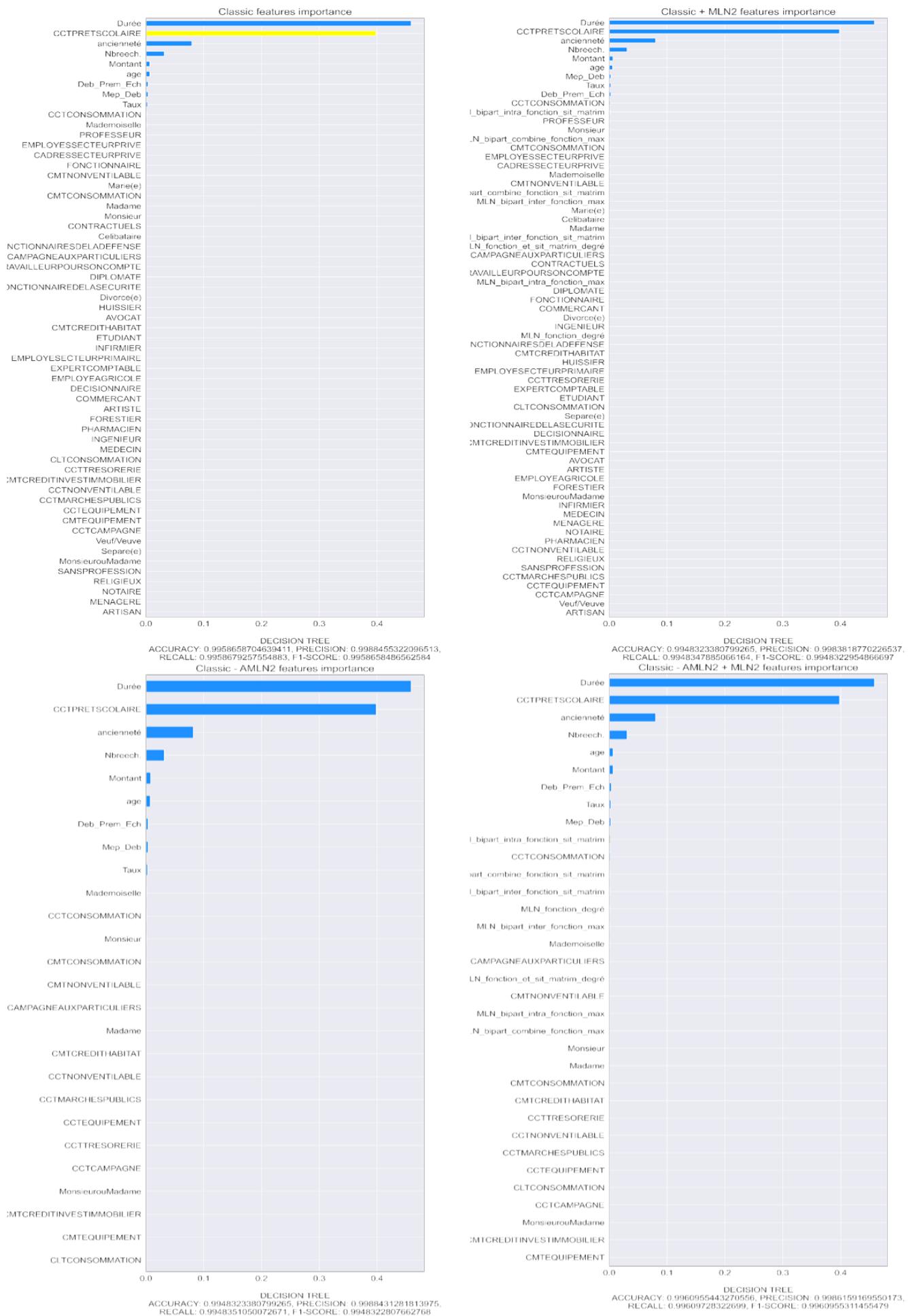


FIGURE 5 – Arbre de Décision (Decision Tree) – meilleur graphe multicouches MLN2

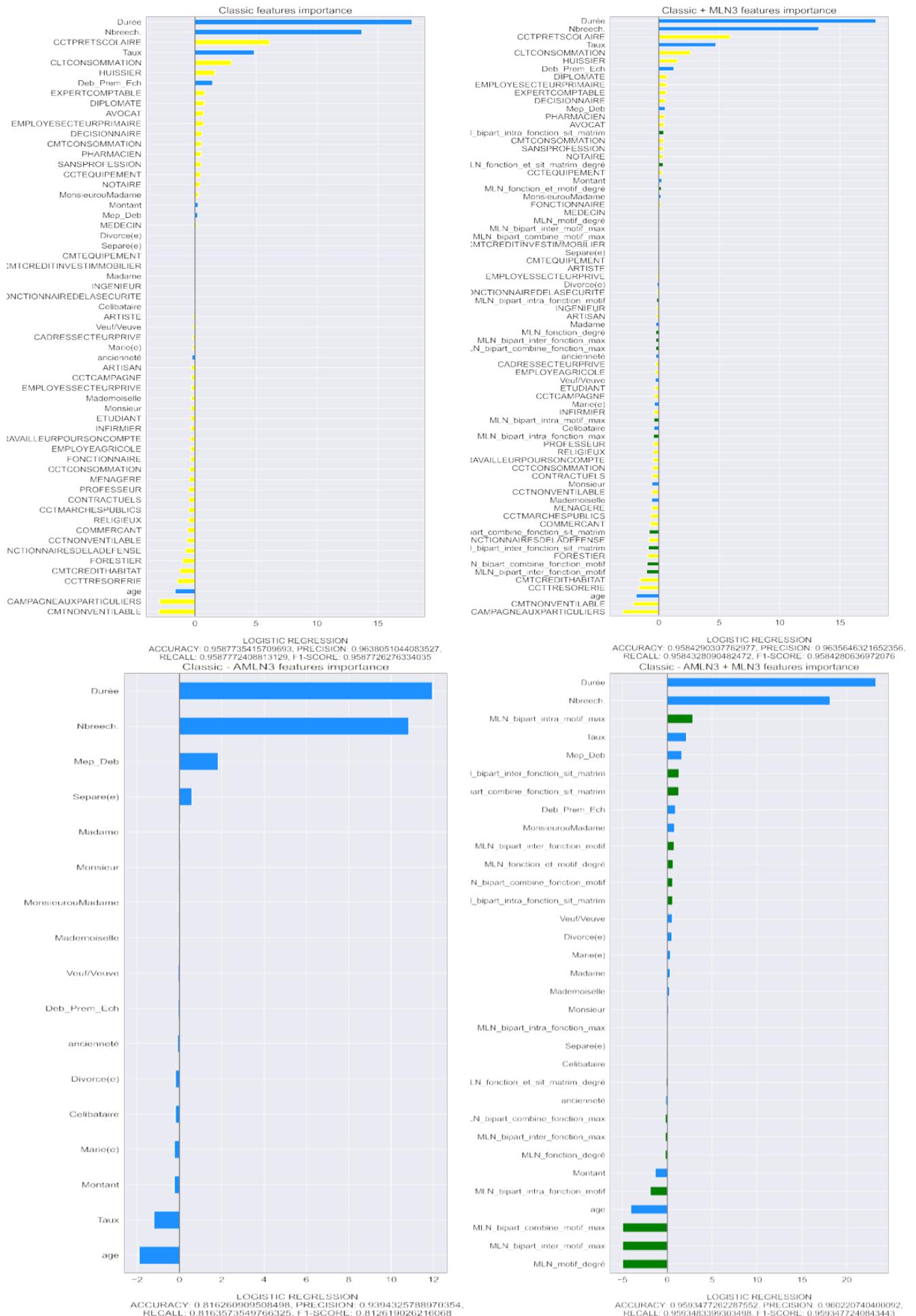


FIGURE 6 – Régression Logistique – meilleur graphe multicouches MLN3

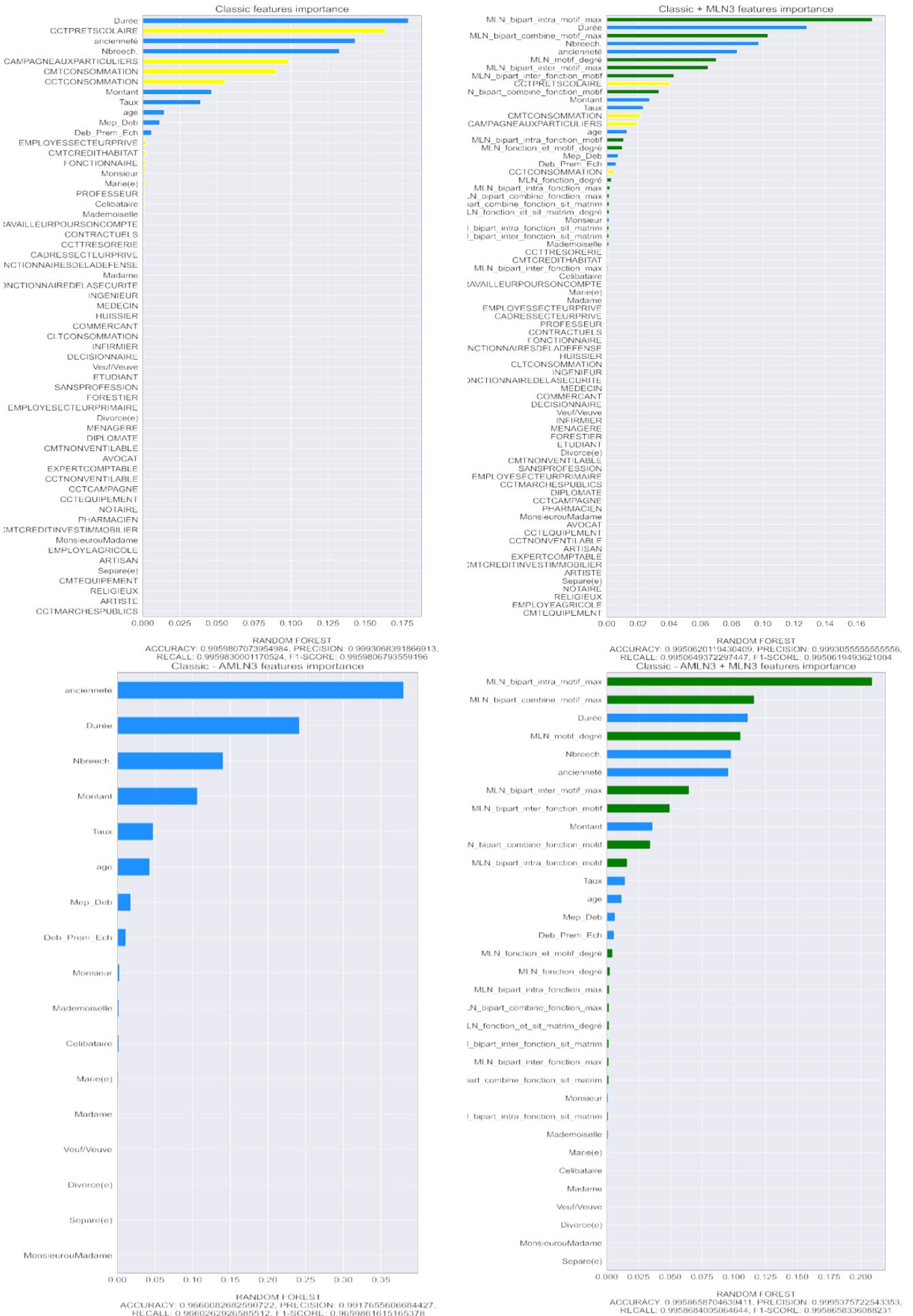


FIGURE 7 – Forêt Aléatoire (Random Forest) – meilleur graphe multicouches MLN3

### 4.3 Descripteurs génériques extraits des graphes multicouches MLNk

Après reproduction des travaux de Óskarsdóttir et al.[4] et compréhension des nouveaux descripteurs, pour simplifier la notation, et pour permettre l'extraction de ces descripteurs dans des graphes multicouches de plus de deux couches, nous proposons de considérer 07 descripteurs génériques à partir des travaux initiaux, et proposons 02 nouveaux descripteurs issus de l'idée de personnalisation du PageRank pour chaque prêt.

Dans la suite du document, **case\_k** représente les noms concaténés par "\_" de 1, 2, ..., k attributs catégoriels du jeu de données, donnant lieu à un MLNk. Par exemple, MLN2 correspond à un graphe multicouches à deux couches qui correspondent à deux attributs catégoriels qui peuvent être "Statut Matrimonial" et "Motif du prêt".

Les 07 attributs génériques considérés pour un MLNk sont les suivants :

- **MLN\_case\_k\_degré** : le nombre de prêts qui sont décrits par les mêmes modalités pour les k attributs du MLNk.
- **MLN\_bipart\_intra\_case\_k** : le score PageRank maximal d'un noeud du prêt parmi ses k représentations dans les k différentes couches du graphe multicouches, lorsque le PageRank est appliqué de manière global à tous prêt suivant la stratégie Intra-Influence présentée dans la section 3.2.
- **MLN\_bipart\_intra\_max\_case\_k** : le score PageRank maximal d'un noeud modalité qui est lié à un des noeuds du prêt parmi ses k représentations dans les k différentes couches du graphe multicouches, lorsque le PageRank est appliqué de manière global à tous prêt suivant la stratégie Intra-Influence présentée dans la section 3.2.
- **MLN\_bipart\_inter\_case\_k** : le score PageRank maximal d'un noeud du prêt parmi ses k représentations dans les k différentes couches du graphe multicouches, lorsque le PageRank est appliqué de manière global à tous prêt suivant la stratégie Inter-Influence présentée dans la section 3.2.
- **MLN\_bipart\_inter\_max\_case\_k** : le score PageRank maximal d'un noeud modalité qui est lié à un des noeuds du prêt parmi ses k représentations dans les k différentes couches du graphe multicouches, lorsque le PageRank est appliqué de manière global à tous prêt suivant la stratégie Inter-Influence présentée dans la section 3.2.
- **MLN\_bipart\_combine\_case\_k** : le score PageRank maximal d'un noeud du prêt parmi ses k représentations dans les k différentes couches du graphe multicouches, lorsque le PageRank est appliqué de manière global à tous prêt suivant la stratégie Combined-Influence présentée dans la section 3.2.
- **MLN\_bipart\_combine\_max\_case\_k** : le score PageRank maximal d'un noeud modalité qui est lié à un des noeuds du prêt parmi ses k représentations dans les k différentes couches du graphe multicouches, lorsque le PageRank est appliqué de manière global à tous prêt suivant la stratégie Combined-Influence présentée dans la section 3.2.

Les 02 nouveaux descripteurs que nous proposons sont issus de l'idée de personnalisation de l'application du PageRank pour chaque prêt. En effet, Óskarsdóttir et al.[4] proposent d'exécuter le PageRank une seule fois sur le graphe multicouches avant d'extraire les nouveaux descripteurs de tous les prêts. Et pourtant l'application du PageRank peut davantage être personnalisé pour chaque prêt, si lors de la personnalisation du PageRank, seuls les noeuds liés au prêt sont activés.

Nous proposons donc les deux nouveaux descripteurs **MLN\_bipart\_ultra\_case\_k** et **MLN\_bipart\_ultra\_max\_case\_k** qui correspondent respectivement aux descripteurs **MLN\_bipart\_combine\_case\_k** et **MLN\_bipart\_combine\_max\_case\_k** lorsque le PageRank est appliqué de façon personnalisée pour chaque prêt. Autrement dit, au lieu d'une exécution PageRank pour tous les prêts, on a autant d'exécution du PageRank que de prêts.

## 4.4 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 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.

Le tableau 3 ci-dessous présentent les comparaisons sur cinq jeux de données (AER, AFB, CREDIT RISK, GERMAN, JAPAN), de l'impact de l'usage des descripteurs issus des graphes multicouches lorsque le PageRank est appliqué de manière globale pour tous les noeuds et lorsque qu'il est appliquée de manière personnalisée pour chaque prêt.

Suivant la métrique *Accuracy*, l'intégration des descripteurs issus du graphe multicouches lorsque le PageRank est appliquée de manière globale permet de faire mieux que le cas classique (uniquement avec les descripteurs du jeu de données) dans 18 cas sur 50 (36%), et le PageRank personnalisé par prêt dans 22 cas sur 50 (44%). Concernant les autres métriques, on a respectivement 14/50 (28%) et 25/50 (50%) suivant la métrique *Precision*, et enfin 15/50 (30%) et 23/50 (46%) suivant les métriques *Rappel* et *F1-score* pour le PageRank global pour tous les noeuds et le PageRank personnalisé par prêt.

Ces résultats montrent que l'usage des descripteurs extraits du graphe multicouches permet d'améliorer les performances des modèles classiques de prédiction du risque de crédit au moins dans 28% des cas si on utilise le PageRank globale pour tous les noeuds et dans au moins 44% des cas si on utilise le PageRank personnalisé par prêt. Ceci confirme l'intérêt de procéder à la personnalisation du PageRank par prêt.

Les pourcentages d'amélioration des techniques classiques de prédiction du risque de crédit par l'usage des descripteurs extraits des graphes multicouches peuvent atteindre jusqu'à 0.3% dans le jeu de données *AFB* de Afriland First Bank, 1% dans le jeu de données *CREDIT RISK*, 4% dans le jeu de données *AER*, 6.5% dans le jeu de données *JAPAN* et enfin 23.4% dans le jeu de données *GERMAN*.

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>		2	<b>2</b>		1	<b>2</b>		1	<b>2</b>

AFB		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	0.9891	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	<b>5</b>		2	<b>5</b>		4	<b>5</b>		4	<b>5</b>

CREDIT_RISK_DATASET		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	0.9263	<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	<b>5</b>		3	<b>7</b>		3	<b>5</b>		3	<b>5</b>

GERMAN		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	0.745	<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	0.755	<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	<b>5</b>		4	<b>5</b>		4	<b>5</b>		4	<b>5</b>

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.					

Après analyse du tableau général des résultats obtenus pour évaluer l'apport de l'intégration des descripteurs extraits des graphes multicouches, nous nous intéressons à la contribution de ces nouveaux descripteurs dans les processus de décision des techniques classiques de prédiction du risque de crédit bancaire. Pour ce faire, nous construisons les diagrammes SHAP des valeurs de Shapley de tous les descripteurs positionnés en entrée des modèles de prédiction.

Les figures 8, 9, 10, 11 et 12 présentent les diagrammes SHAP des Top-20 des valeurs de shapley des descripteurs des modèles de prédiction de type **Classic + MLN** respectivement basé sur *SVM*, *Régression Logistique*, *Arbre de Décision*, *Fôret d'Arbres de Décision* et *XGBoost* dans les jeux de données AFB, CREDIT\_RISK, GERMAN et JAPAN. Ceux du jeu de données AER n'y figurent pas par souci d'espace et de présentation du document.

Dans les 5 figures citées, nous avons 20 diagrammes SHAP sur lesquels les barres de couleur verte correspondent aux attributs des graphes multicouches. On constate que 14/20 (70%) diagrammes ont une présence significative des barres vertes, 02/20 (10%) diagrammes ont une seule barre verte et les 04/20 (20%) restant ne contiennent pas du tout de barre verte. Ceci illustre la pertinence des attributs extraits des graphes multicouches.

Lorsqu'on s'intéressant au meilleur type de descripteurs extraits des graphes multicouches, c'est-à-dire, le descripteur associé à la plus longue barre verte de chacun des 16 diagrammes qui contiennent des barres vertes, on constate que dans 9/16 (56%) des cas c'est un descripteur que nous proposons **MLN\_bipart\_ultra\_max**, dans 4/16 (25%) des cas c'est encore un descripteur que nous proposons **MLN\_bipart\_ultra** et dans seulement 3/16 (19%) des cas se sont les 07 autres descripteurs proposés par Óskarsdóttir et al.[4]. Soit un total de 13/16 (81%) pour les attributs issus de l'application du PageRank personnalisé par prêt.

Il est donc nécessaire de poursuivre l'exploration de l'idée de personnalisation de l'application du PageRank par prêt pour les deux autres stratégies INTRA et INTER décrites dans la section 3.2 pour l'extraction des caractéristiques à partir du graphe multicouches construit.

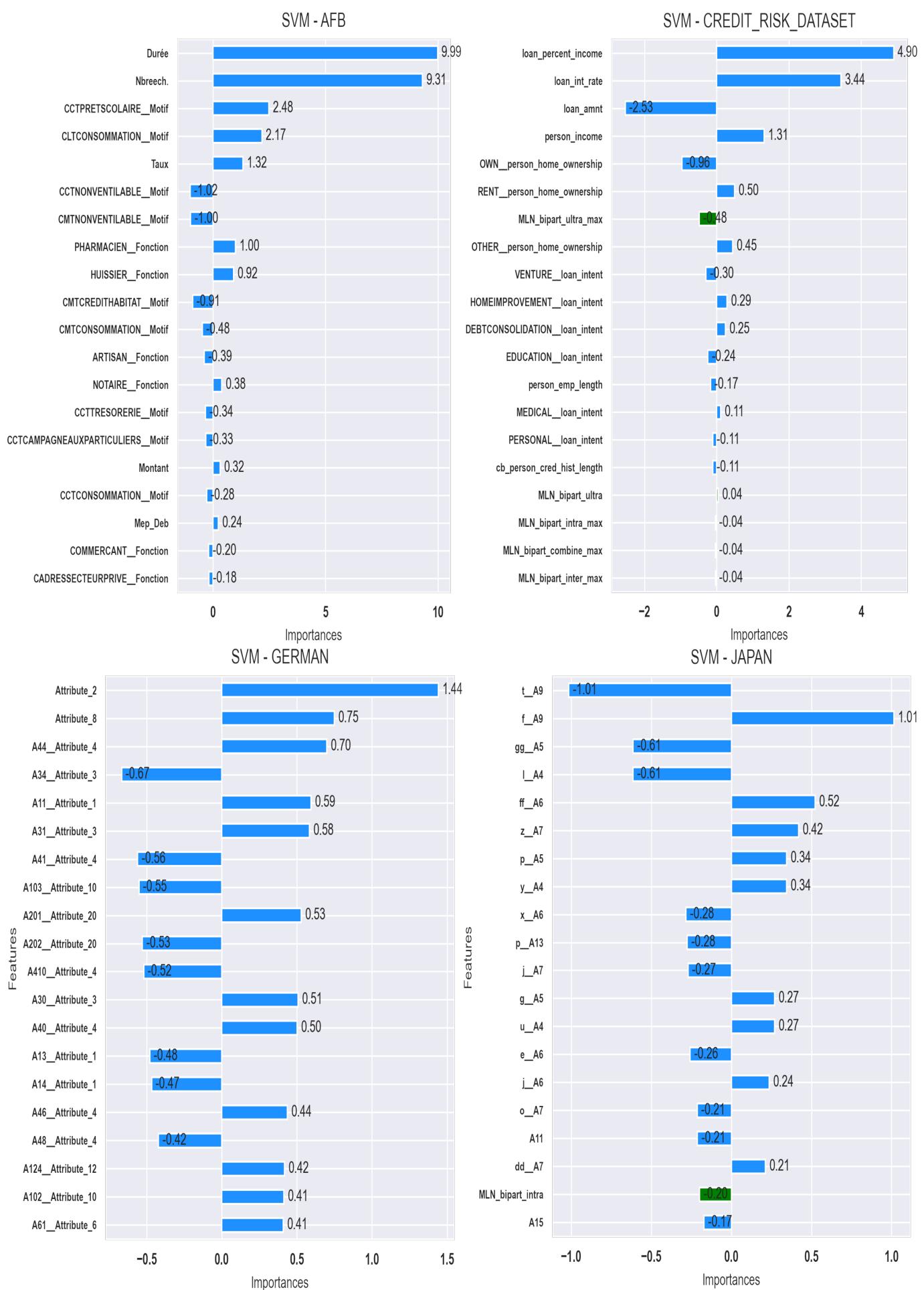


FIGURE 8 – Diagramme SHAP de l’importance des attributs pour le modèle SVM

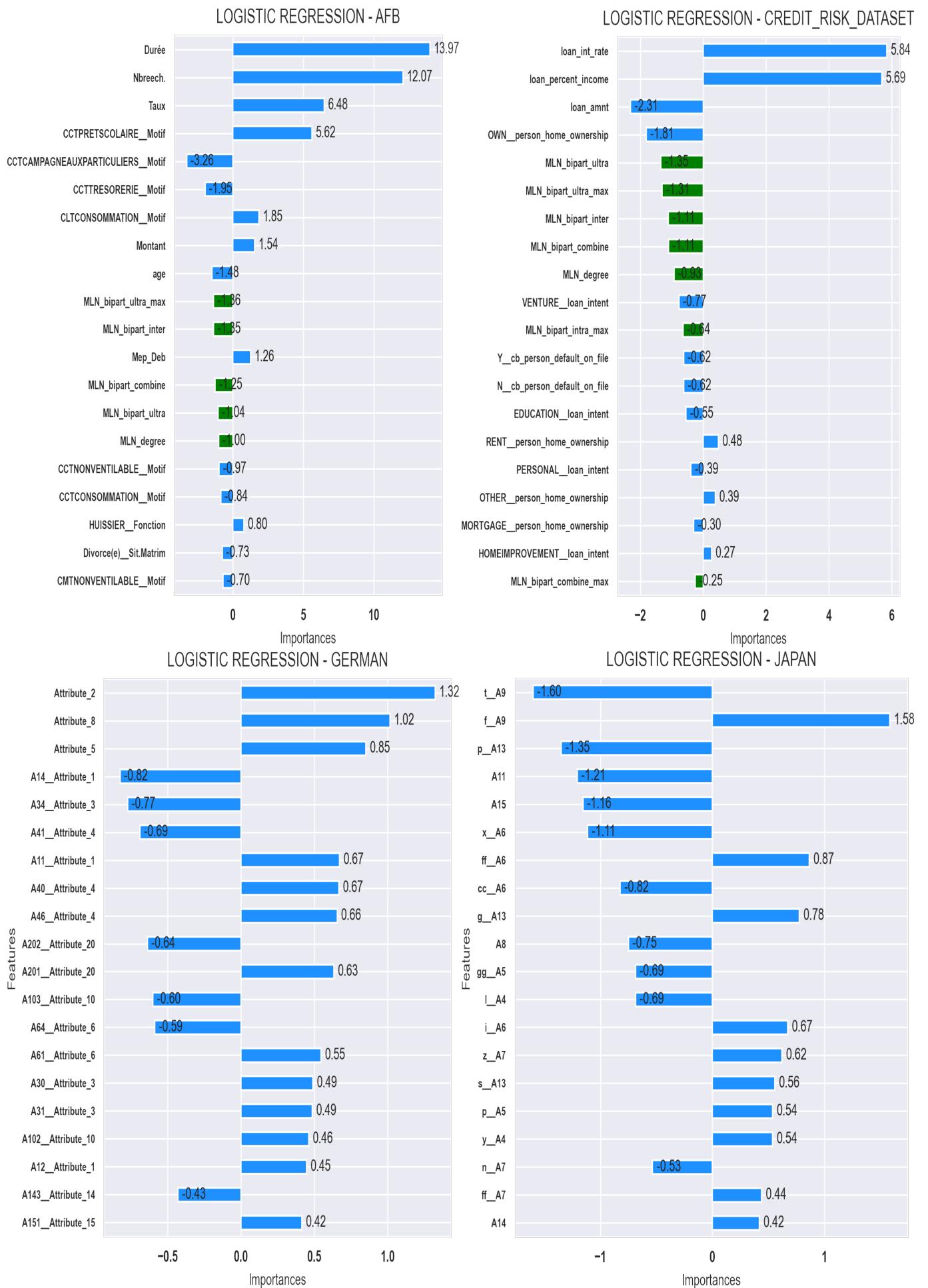


FIGURE 9 – Diagramme SHAP de l’importance des attributs pour la Régression Logistique

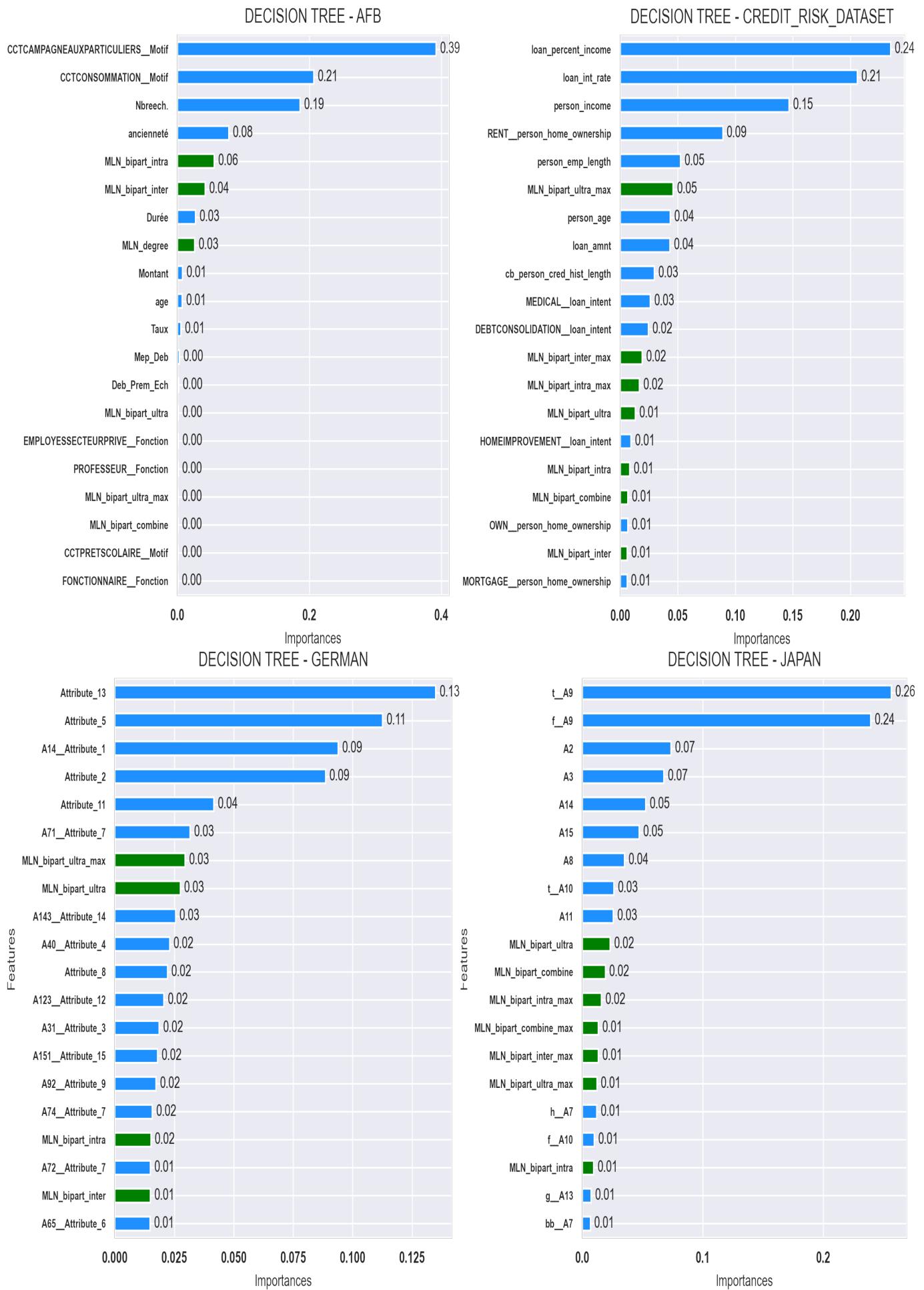


FIGURE 10 – Diagramme SHAP de l'importance des attributs pour l'Arbre de Décision

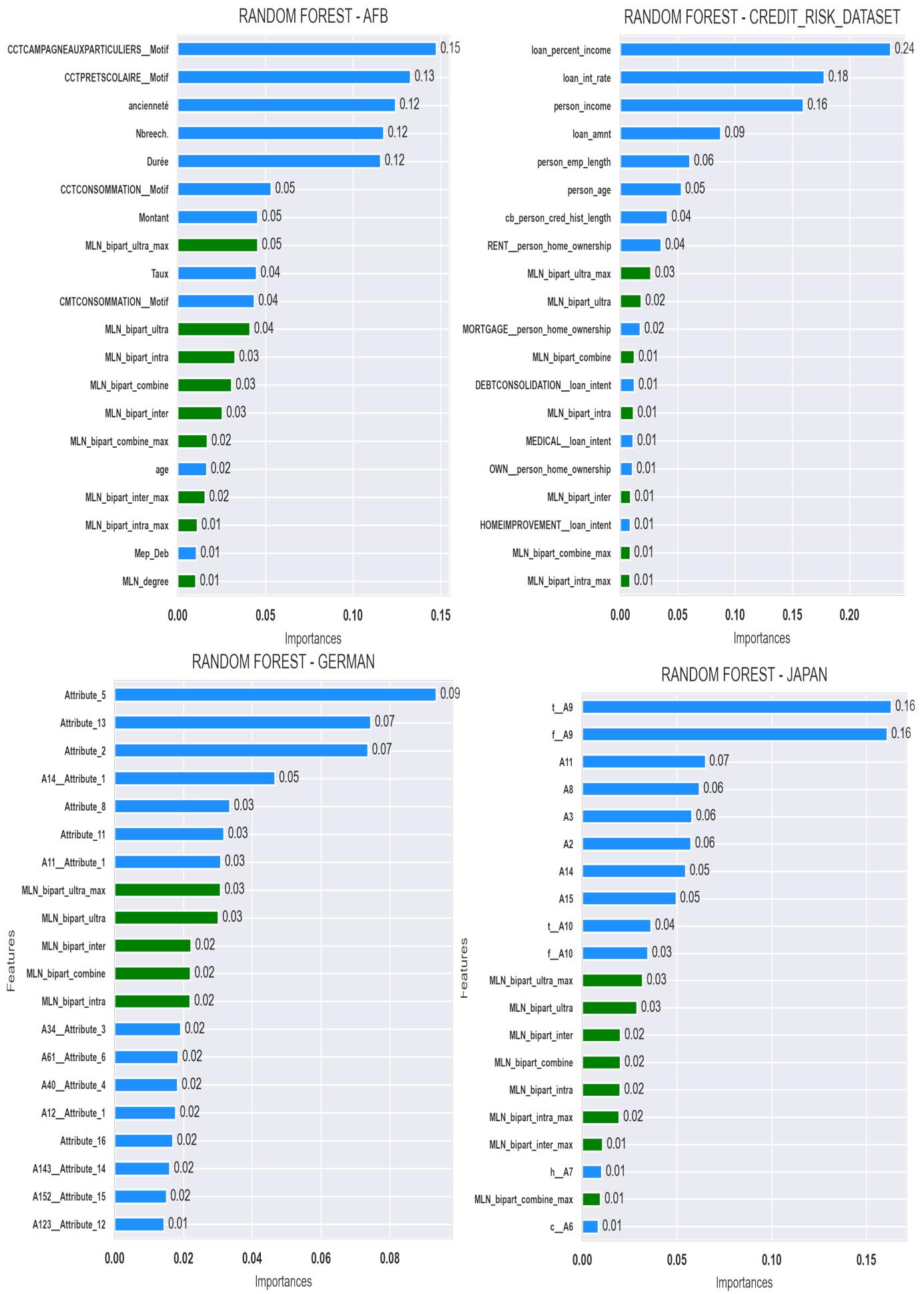


FIGURE 11 – Diagramme SHAP de l’importance des attributs pour la Fôret Aléatoire

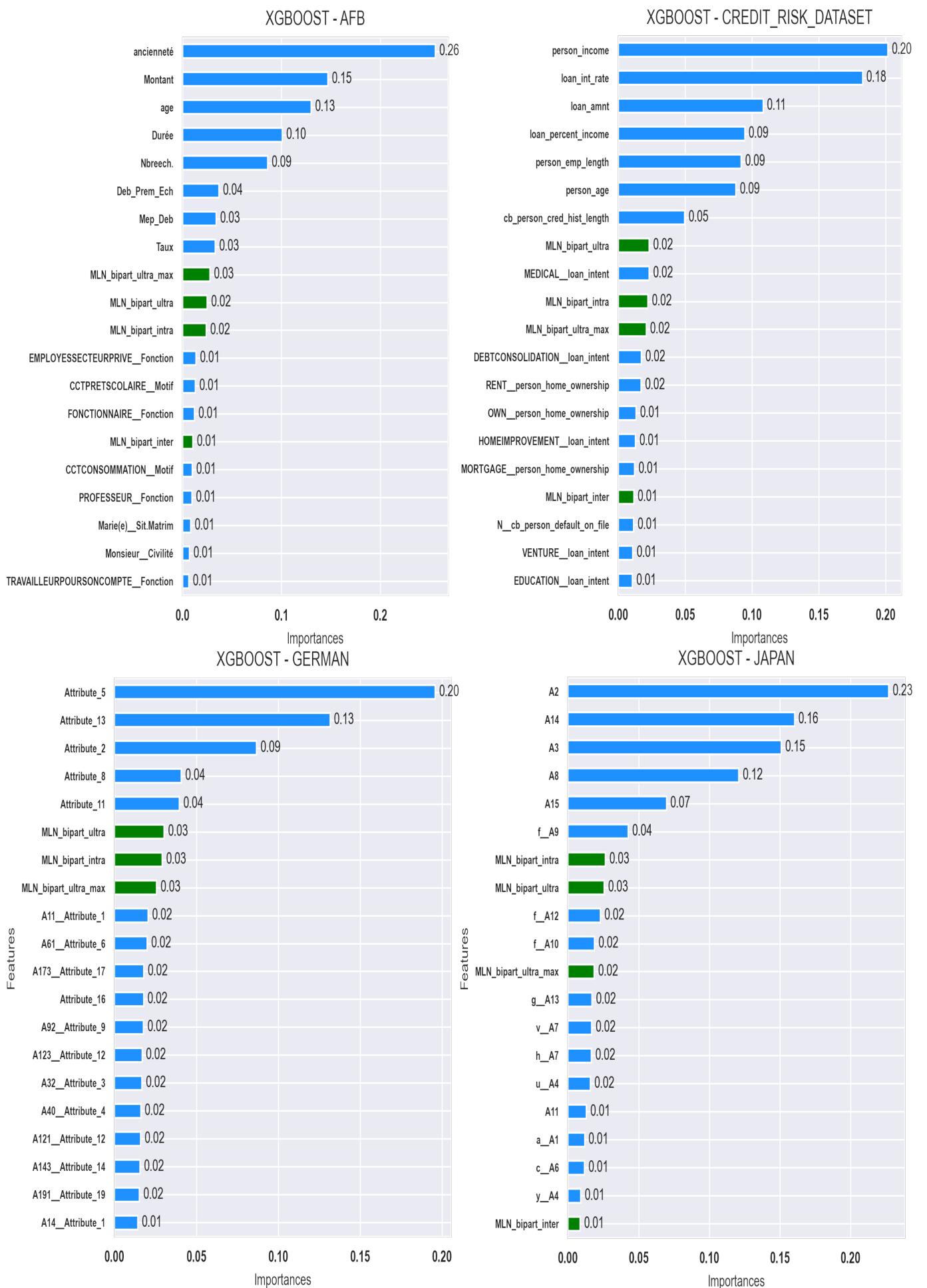


FIGURE 12 – Diagramme SHAP de l’importance des attributs pour XGBoost

	SVM	LR	DT	RF	XGB	TOTAL
AER (13 + 9)	8	7	7	8	7	37
AFB (59 + 9)	0	5	6	9	4	24
CREDIT RISK (19 + 9)	5	7	7	7	4	30
GERMAN (61 + 9)	0	0	4	5	3	12
JAPAN (46 + 9)	1	0	7	8	4	20
<b>TOTAL</b>	14	19	31	37	22	

TABLE 4 – Nombre d’attributs des graphes multicouches qui sont présents dans le Top-20 des attributs qui contribuent le plus aux décisions des modèles de prédiction dans les différents jeux de données. Entre parenthèse, c’est le nombre total d’attributs du jeu de données associé.

Le tableau 4 présente un état du nombre de descripteurs du graphe multicouches qui sont présents dans le Top-20 des descripteurs qui contribuent le plus aux décisions des modèles de prédiction. Sur les 45 places de classement possibles que peuvent occuper les attributs extraits des graphes multicouches en considérant tous les jeux de données et tous les modèles, nous faisons les constats suivants :

- Dans le jeu de données AER, les attributs des graphes multicouches occupent 37/45 (82%) des places possibles, et on a 24/45 (53%), 30/45 (67%), 12/45 (27%), 20/45 (44%) respectivement dans les jeux de données AFB, CREDIT\_RISK, GERMAN et JAPAN.
- En s’attardant sur les modèles de prédiction, dans le cas du SVM, les attributs des graphes multicouches occupent 14/45 (31%) des places possibles, et on a 19/45 (42%), 31/45 (69%), 37/45 (82%) et 22/45 (49%) respectivement pour les modèles Régression Logistique (LR), Arbre de décision (DT), Forêt Aléatoire (RF) et XGBoost (XGB).

Des constats précédents, nous pouvant faire les analyses suivantes :

- Les attributs des graphes multicouches ne sont pas très intéressants pour les jeux de données GERMAN et JAPAN car la plupart ne sont pas dans le Top-20. Par contre, ces attributs sont intéressants pour les jeux de données AFB et CREDIT\_RISK car la plupart sont dans le Top-20, mais nous ne pouvons pas jugé le cas de AER, car le nombre total d’attributs est à peine de 22 (et nous observons le Top-20).
- Les attributs des graphes multicouches semblent ne pas être intéressants pour le modèle SVM, semblent assez intéressants pour la Régression Logistique et vraiment intéressants pour les modèles qui reposent sur les Arbres de Décision (DT, RF et XGBoost).

## 4.5 Choix des attributs à considérer dans le graphe multicouches

Dans l’article de Óskarsdóttir et al.[4] que nous considérons comme base de l’utilisation des graphes multicouches pour la prédiction du risque de crédit, les auteurs implémentent uniquement un graphe multicouches à 2 couches, cependant, une bonne compréhension des descripteurs qu’ils proposent permet d’envisager la construction des graphes multicouches à  $k$  couches, avec  $k \in \{1, 2, \dots, N\}$  et  $N$  le nombre d’attributs catégoriels du jeu de données.

Une limite directe à leurs travaux est le fait qu’ils n’exploront pas la possibilité d’un graphe multicouches avec autant de couches que d’attributs catégoriels, et surtout qu’ils ne proposent pas un moyen de choisir les attributs catégoriels à considérer comme couches du graphe multicouches. C’est pour ces raisons que dans cette section, nous explorons la possibilité de construire des graphes multicouches avec des nombres de couches différents, évaluons l’impact du nombre de couches sur les performances des modèles de prédiction, et expérimentons une piste pour la sélection des attributs à considérer dans les graphes multicouches d’au moins deux couches.

#### 4.5.1 Graphes multicouches avec des nombres différents de couches

En tenant compte de la présentation des descripteurs génériques dans la section 4.3, nous pouvons construire des graphes multicouches à une seule couche MLN1 (avec un seul attribut catégoriel comme couches), mais aussi des graphes multicouches avec deux couches MLN2 (avec deux attributs catégoriels) et même  $N$  couches MLN\_All (avec tous les attributs catégoriels comme couches). Les résultats des expérimentations effectuées pour la prédiction du risque de crédit bancaire sur les 5 jeux de données considérés en utilisant les graphes multicouches MLN1, MLN2 et MLN\_All sont présentés dans les tableaux 5 et 6.

Dans chaque jeu de données, la ligne **Total** contient le nombre de fois que le graphe MLN $k$  en colonne a la meilleure performance comparé aux autres (on ignore les cas où il y a égalité de performances entre les trois MLN $k$  considérés – MLN1, MLN2 et MLN\_All). En effectuant les sommes des totaux de tous les jeux de données, pour chaque MLN $k$  et suivant chaque métrique, on a Accuracy (27, 48, 8), Précision (13, 53, 5), Rappel (16, 57, 5) et enfin F1-score (21, 52, 6) respectivement pour les graphes MLN $k$  (MLN1, MLN2, MLN\_All). Les pourcentages des cas où chaque MLN $k$  est meilleur comparé aux autres suivant les 04 métriques considérées sont :

- **Accuracy** : MLN1 27/83 (33%), MLN2 48/83 (58%) et MLN\_All 8/83 (10%)
- **Précision** : MLN1 13/71 (18%), MLN2 53/71 (75%) et MLN\_All 5/71 (07%)
- **Rappel** : MLN1 16/78 (21%), MLN2 57/78 (73%) et MLN\_All 5/78 (06%)
- **F1-score** : MLN1 21/79 (27%), MLN2 52/79 (66%) et MLN\_All 6/79 (08%)

On remarque que quelque soit la métrique, l'ordre entre les MLN $k$  est le même : MLN2 est toujours le meilleur, suivi de très loin par MLN1 et MLN\_All est toujours en dernière position dans le classement. Ce qui permet de déduire qu'il est important d'utiliser des graphes multicouches avec plusieurs couches (plus de un couche), cependant il est déconseillé d'utiliser d'office tous les attributs catégoriels possible, et donc il est nécessaire de pouvoir sélectionner les attributs qui vont être considérés comme couche du graphe multicouches.

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^	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	0.9588	0.0	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	0.9588	0.0	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	0.9588	0.0	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	0.9588	0.0	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	0.9637	0.0	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	0.9637	0.0	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	0.9637	-0.56	0.0	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	0.8643	0.0	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	0.8643	0.0	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	0.9641	0.0	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	0.9641	0.0	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	0.9641	0.0	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	0.9641	0.0	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	0.8621	0.0	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	0.8621	0.0	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	4		1	4		1	4		1	4		1	4	

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>	5	2	6	<b>10</b>	1	6	<b>11</b>	1	8	7	2		

CREDIT_RISK			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

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

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	0.913	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	0.913	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	0.0	<b>1.41</b>	0.12	0.12	0.8794	0.93	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	0.9242	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	0.9242	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

TABLE 6 – Comparaison des taux d'amélioration de la performance du cas classique (Classic) par les graphes multicouches, MLN1, MLN2 et MLN\_All dans GERMAN et JAPAN.

#### 4.5.2 Sélection des attributs à considérer dans un graphe multicouches

Suite aux conclusions de la sous-section précédente, nous entamons une piste qui repose sur l'utilisation de MLN1 pour la sélection des attributs à considérer pour construire les graphes multicouches d'au moins deux couches. En effet, la performance en terme de qualité de prédiction d'un graphe MLN1 qui considère un seul attribut catégoriel  $A$  comme couche du graphe multicouches peut être considérée comme un élément de justification ou non, de considérer l'attribut  $A$  concerné comme bon pour la construction des graphes multicouches.

Pour vérifier cette hypothèse nous procérons comme suit :

1. Construire le graphe MLN1 pour chaque attribut catégoriel  $A$
2. Évaluer leurs performances suivant les métriques Accuracy, Précision, Rappel et F1-score.
3. Comparer l'impact de la version numérique de  $A$  résultante du MLN1 à sa version originale catégorielle ; ça revient à comparer les résultats de **Classic** et **Classic+MLN1-Att**.

4. Lorsque la performance de **Classic+MLN1-Att** est meilleure que celle de **Classic**, on considère que la version numérique de *A* est meilleure et on l'étiquette **Good MLN1**. Dans le cas contraire, on l'étiquette **Bad MLN1**.
5. Construire les graphes multicouches MLN2 à deux couches avec tous les couples possibles d'étiquettes d'attributs (**Good MLN1, Good MLN1**), (**Good MLN1, Bad MLN1**) et (**Good MLN1, Bad MLN1**).
6. Comparer les résultats des graphes multicouches MLN2 de chaque couple à ceux de la prédiction de risque de crédit avec les attributs classiques du jeu de donnée (**Classic**).
7. Identifier le couple d'étiquette qui permet d'avoir les meilleurs résultats.
8. Conclure sur validité de notre hypothèse.

Le tableau 7 présente la comparaison des performances des graphes multicouches MLN2 construits à partir des couples d'attributs d'étiquette (**Good MLN1, Good MLN1**), (**Good MLN1, Bad MLN1**) et (**Good MLN1, Bad MLN1**) dans tous les jeux de données et suivant les métriques Accuracy, Précision et F1-score.

Les colonnes "**Classic >**" contiennent le nombre de fois que la performance du MLN2 du couple étiqueté associé est pire que celle de (**Classic**) ; Les colonnes "**Classic =**" contiennent le nombre de fois que la performance du MLN2 du couple étiqueté associé est égale à celle de (**Classic**) ; et enfin les colonnes "**Classic <**" contiennent le nombre de fois que la performance du MLN2 du couple étiqueté associé est supérieure à celle de (**Classic**).

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

TABLE 7 – Comparaison des performances des graphes multicouches MLN2 construits à partir des couples d'attributs d'étiquette (Good MLN1, Good MLN1), (Good MLN1, Bad MLN1) et (Good MLN1, Bad MLN1) dans tous les jeux de données.

Lorsqu'on résume les valeurs dans le tableau 7, on constate que :

- Les couples d'attributs d'étiquette (**Good MLN1, Good MLN1**) sont pires que (**Classic**) dans 436/2397 (18%) des cas, sont d'égal performance que (**Classic**) dans 126/2387 (5%) des cas et sont meilleure que (**Classic**) dans 1835/2387 (77%) des cas.
- Les couples d'attributs d'étiquette (**Good MLN1, Bad MLN1**) sont pires que (**Classic**) dans 971/2445 (40%) des cas, sont d'égal performance que (**Classic**) dans 243/2445 (10%) des cas, et sont meilleure que (**Classic**) dans 1231/2445 (50%) des cas.

- Les couples d'attributs d'étiquette (**Bad MLN1**, **Bad MLN1**) sont pires que (**Classic**) dans 2175/5078 (43%) des cas, sont d'égal performance que (**Classic**) dans 1876/5078 (37%) des cas, et sont meilleure que (**Classic**) dans 1027/5078 (20%) des cas.

On remarque que lorsque chacun des deux attributs du graphe multicouches MLN2 sont de type (**Good MLN1**), les performances du graphe multicouches MLN2 résultant sont meilleures que ceux de l'exécution classique des modèles de prédiction (**Classic**) dans 77% des cas possibles. De plus, dès l'instant qu'on intègre dans le couple des attributs du MLN2, un qui est étiqueté (**Bad MLN1**), le graphe résultant est meilleur que (**Classic**) dans 50% des cas possibles. En fin, si les deux attributs du MLN2 sont étiqueté (**Bad MLN1**) le graphe résultant est meilleur que (**Classic**) dans seulement 20% des cas possibles.

Sur la base des remarques faites, l'hypothèse selon laquelle les attributs dont le graphe MLN1 associé permet d'améliorer le cas classique sont de bons candidats à sélectionner pour des graphes multicouches d'au moins deux couches. On peut donc proposer le protocole suivant pour le choix des k attributs à considérer pour les graphes multicouches à plusieurs couches :

1. Considérer tous les attributs catégoriels du jeu de données.
2. Évaluer les performances des graphes MLN1 de chaque attribut.
3. Sélectionner les k attributs associés aux k MLN1 qui ont les k plus grands taux d'amélioration de la version classique de manière positive suivant la métrique cible.

## 4.6 Configurations qui permettent d'avoir les meilleures performances

Sur la base des résultats des sections 4.4 et 4.5, nous décrivons les configurations les plus susceptibles d'entrainer de bons résultats. C'est à cet effet que la figure 13 illustre les contributions des types de modèles de prédiction de base (SVM, Régression Logistique, Arbre de Décision, Forêt Aléatoire d'arbres de décision et XGBoost), les types de lots d'attributs en entrée (Classic, Classic-Att, Classic+MLN, Classic+MLN-Att), le type de personnalisation du PageRank (Globale pour tous les prêts, Personnalisée par prêt) et enfin le nombre de couches à considérer dans le graphe multicouches (MLN1, MLN2, MLN\_All).

A partir des graphiques de la figure 13, on peut arrêter les conclusions suivantes :

- **XGboost** est le modèle de base qui permet généralement d'avoir les meilleures résultats.
- **Classic+MLN** est le type de lots d'attributs qui permet généralement d'avoir les meilleures performances. Ce qui confirme l'importance des descripteurs du graphe multicouches.
- **Le PageRank personnalisé par prêt** permet généralement d'avoir les meilleures performances. Ce qui renforce la perspective de personnaliser par prêt, les stratégies INTRA et INTER de la section 3.2 pour l'extraction des descripteurs du graphe multicouches.
- **Les graphes multicouches MLN2 à deux couches** permettent généralement d'avoir les meilleures performances parmi MLN1, MLN2 et MLN\_All. Ce qui confirme la nécessité de sélectionner des attributs à considérer pour la construction des graphes multicouches.

Et donc pour augmenter les chances d'avoir les meilleures performances en terme de qualité de prédiction du risque de crédit, à la fin de cette étude, on recommande d'appliquer le modèle **XGboost** sur le lot des descripteurs de type **Classic+MLN** extraits en utilisant le **PageRank personnalisé par prêt** appliqué sur un **graphe multicouches MLN2 à deux couches**.

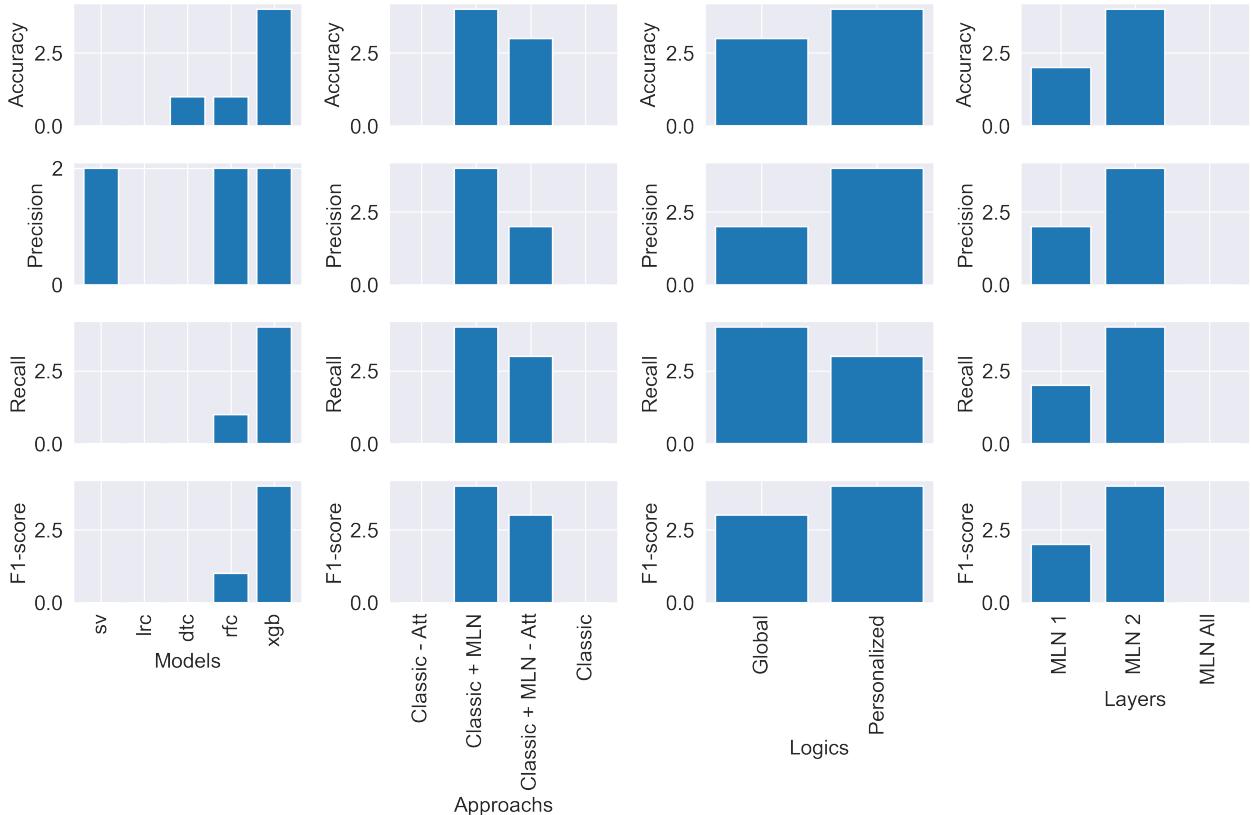


Figure 13 – Configurations qui permettent d'avoir les meilleures performances.

## 5 Conclusion et perspectives

### 5.1 Conclusion

L'objectif de ce travail était d'explorer les voies d'utilisation de la modélisation des données bancaires par des graphes afin d'améliorer la qualité des modèles classiques de prédiction du risque de crédit. Nous nous sommes attardé sur le cas des graphes multicouches dont l'usage nécessite d'abord la modélisation du graphe en considérant des attributs catégoriels comme couches de ces graphes, ensuite l'application du PageRank pour extraire des caractéristiques qui sont ajoutées en entrée du processus d'apprentissage des modèles de prédiction.

La compréhension des travaux existants nous a permis de les répliquer sur un jeu de données d'Afriland First Bank et sur quatre autres jeux de données standards dédiés à la prédiction du risque de crédit. Cette première vague d'expérimentation a servie à confirmer l'importance des descripteurs extraits des graphes multicouches.

Par ailleurs, nous avons abordé des pistes de solutions aux limites des travaux existants notamment en proposant de Personnaliser le PageRank pour chaque prêt au lieu d'avoir une seule exécution PageRank pour tous les prêts. La pertinence de cette première proposition s'illustre par le fait dans 81% des cas, l'attribut le plus important parmi ceux extraits du graphe multicouches est issu de la Personnalisation du PageRank pour chaque prêt.

De plus, nous avons évalué l'impact du nombre d'attributs considérés comme couches dans les graphes multicouches, et il en ressort que ni les graphes multicouches associé un seul attribut, ni le graphe multicouches associé à tous les attributs catégoriels ne permet d'avoir généralement les meilleures prédictions du risque de crédit, mais c'est le graphe multicouches à deux couches qui le permet en général. Le constat précédent a confirmé la nécessité de procéder à une sélection des attributs à considérer dans le graphe multicouche, contrairement aux travaux existants qui

n'en propose aucune. Ainsi, nous avons proposé un protocole pour le choix des attributs de construction des graphes multicouches à  $k$  couches, dont le test sur les graphes multicouches à deux couches a été un succès à plus de 75%.

## 5.2 Perspectives

Rendu à la fin des travaux présentés jusqu'à cet instant et compte tenu des limites actuelles de la littérature concernant la prédiction du risque de crédit bancaire en utilisant les attributs extraits des graphes multicouches, nous envisageons les perspectives suivantes :

- Étendre l'application de la personnalisation du PageRank par prêt aux stratégies d'extraction des caractéristiques Intra et Inter proposés dans la littérature actuelle.
- Appliquer le protocole proposé pour la construction des graphes multicouches de plus de deux couches d'attributs.
- Proposer une stratégie de construction du graphe multicouches à partir de tous les types d'attributs, y compris les attributs quantitatifs.
- Intégrer dans les graphes multicouches une dimension associée aux deux classes des prêts dans le phase d'apprentissage des modèles de prédiction. L'intégration de telles informations dans la phase d'apprentissage peut améliorer capacités des descripteurs à discriminer les prêts de ces différentes classe.
- Proposer de nouvelles stratégies d'exploitation des graphes multicouches autres que le PageRank pour l'extraction de nouvelles caractéristiques.

## Références

- [1] H. Ince and B. Aktan, “A comparison of data mining techniques for credit scoring in banking : A managerial perspective,” *Journal of Business Economics and Management*, vol. 10, no. 3, pp. 233–240, 2009.
- [2] S. K. Trivedi, “A study on credit scoring modeling with different feature selection and machine learning approaches,” *Technology in Society*, vol. 63, p. 101413, 2020.
- [3] J. Nalić, G. Martinović, and D. Žagar, “New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers,” *Advanced Engineering Informatics*, vol. 45, p. 101130, 2020.
- [4] M. Oskarsdottir and C. Bravo, “Multilayer network analysis for improved credit risk prediction,” *Omega*, vol. 105, p. 102520, 2021.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] P. Giudici, B. Hadji-Misheva, and A. Spelta, “Network based credit risk models,” *Quality Engineering*, vol. 32, no. 2, pp. 199–211, 2020.