The background of the slide is a collage of business-related images. At the top, there's a photo of people's hands pointing at a document with a pie chart. Below this, on the right, is a large dark grey triangle with a teal triangle inside it. At the bottom left, there's a teal triangle containing a bar chart and a line graph. The main title is centered on the left side.

# **PREDICTION DE L'ÉLIGIBILITE AUX PRETS BANCAIRES A L'AIDE DES MODELES MACHINE LEARNING ET ANALYSE COMPARATIVE**

*Alloumi Saad  
Belabbas Youcef*

## Table des abréviations :

Abréviations	Signification
AUC	L'aire sous la courbe ROC
SVM	Machine à vecteurs de support
DM	Dépenses mensuelles en euro
RM	Revenu annuel en millier d'euro
FAM	Nombre d'individus dans la famille
EL_Oui	Épargne logement
EXP	Nombre d'année d'expérience professionnel
Educ	Niveau d'éducation

## Listes des figures :

- 1- Boxplot de la relation entre prêt personnel et l'âge
- 2- Boxplot de la relation entre prêt personnel et l'expérience
- 3- Boxplot de la relation entre le prêt personnel et le nombre d'individus par famille
- 4- Diagramme en barres de la relation entre le prêt personnel et l'éducation
- 5- Boxplot de la relation entre prêt personnel et le revenu
- 6- Boxplot de la relation entre le prêt personnel et les dépenses mensuelles en euro
- 7- Diagramme en barres de la relation entre prêt personnel et épargne logement
- 8- Diagramme en barres des rapports de cote
- 9- Arbre de décision
- 10- Diagramme des barres qui représente l'importance des variables explicatives
- 11- Diagramme en barres qui représente l'indice AUC selon chaque model
- 12- Diagramme en barres qui représente une comparaison d'AUC avec et sans agrégation entre les différents models
- 13- Diagramme en barres qui représente une comparaison du Rappel avec et sans agrégation entre les différents modèles
- 14- Diagramme en barres représentant la prédiction de l'octroi du prêt



# Sommaire

## Table des matières

<b>Introduction.....</b>	<b>3</b>
<b>Partie 1 : Profil Socio-économique de nos Clients : Exploration des Facteurs Clés d'Accord de Prêt Personnel.....</b>	<b>4</b>
<b>1-1 Profil démographique de nos clients.....</b>	<b>4</b>
<b>1-2 Profil financier de nos clients : .....</b>	<b>9</b>
<b>Partie 2 : Évaluation Préliminaire de l'Importance des Variables .....</b>	<b>12</b>
<b>2-1 Modèle Probit : .....</b>	<b>12</b>
<b>2-2 Arbre de décision : .....</b>	<b>14</b>
<b>Partie 3 : MISE EN COMPETITION DES MODELES ET VALIDATION .....</b>	<b>16</b>
<b>Partie 4 : Prédiction de l’octroi du prêt avec les nouvelles données .....</b>	<b>21</b>
<b>Conclusion : .....</b>	<b>22</b>

## Introduction

Ce rapport s'inscrit dans une démarche analytique visant à éclairer les décisions stratégiques au sein de notre institution financière. Plus spécifiquement, cette étude a pour objectif d'analyser les déterminants de l'accord de prêt personnel. Conduite dans le cadre de notre stratégie de gestion des risques, cette analyse vise à fournir des éléments précieux pour faciliter la prise de décision des responsables.

La première section de ce rapport se concentrera sur une exploration approfondie du contexte sociodémographique de notre clientèle. En analysant des variables clés telles que l'âge, l'expérience, le revenu, la taille de la famille, les dépenses mensuelles moyennes, le niveau d'éducation, la détention d'un compte épargne logement, et l'accord du prêt personnel, nous chercherons à identifier des corrélations significatives. Cette démarche analytique vise à anticiper et à comprendre les comportements liés à la décision d'accorder ou de refuser le prêt, renforçant ainsi notre capacité à gérer les risques associés à cette activité financière.

Dans la deuxième section, nous explorerons l'importance des variables à travers un ou deux modèles, mettant en lumière les facteurs déterminants qui influent sur la probabilité d'accord du prêt. Ensuite, dans la troisième section, nous adopterons une approche d'apprentissage automatique (ML) pour résoudre un problème de classification supervisée. Nous entraînerons plusieurs modèles ML afin de sélectionner celui qui offre les performances optimales. La dernière phase de cette section consistera à évaluer la robustesse du modèle sélectionné en le testant sur de nouvelles données. Cela nous permettra de vérifier sa capacité à prédire efficacement les clients qui obtiendront un prêt et ceux qui ne l'obtiendront pas.

Cette méthodologie complète est conçue pour offrir des informations stratégiques solides et faciliter la prise de décision des acteurs clés au sein de notre institution financière.

## Partie 1 : Profil Socio-économique de nos Clients : Exploration des Facteurs Clés d'Accord de Prêt Personnel

Dans cette section, nous explorons le profil socio-économique de nos clients pour mieux comprendre les déterminants de l'accord de prêt personnel. En analysant des variables telles que l'âge, le revenu, la taille de la famille, et d'autres, nous visons à établir des liens significatifs.

### 1-1 Profil démographique de nos clients

#### ❖ Liaison entre prêt personnel et l'âge de notre échantillon de clients

L'âge peut influencer la décision d'obtenir un prêt personnel, car il est souvent associé à des étapes importantes de la vie, telles que le début de la carrière professionnelle ou la planification familiale, ce qui peut impacter les besoins financiers et la propension à obtenir un prêt.

Prêt Personnel	Count	Mean	Std	Min	25%	50%	75%	Max
Non	4415	45.36	11.46	23	35	45	55	67
Oui	479	45.10	11.59	26	35	45	55	65

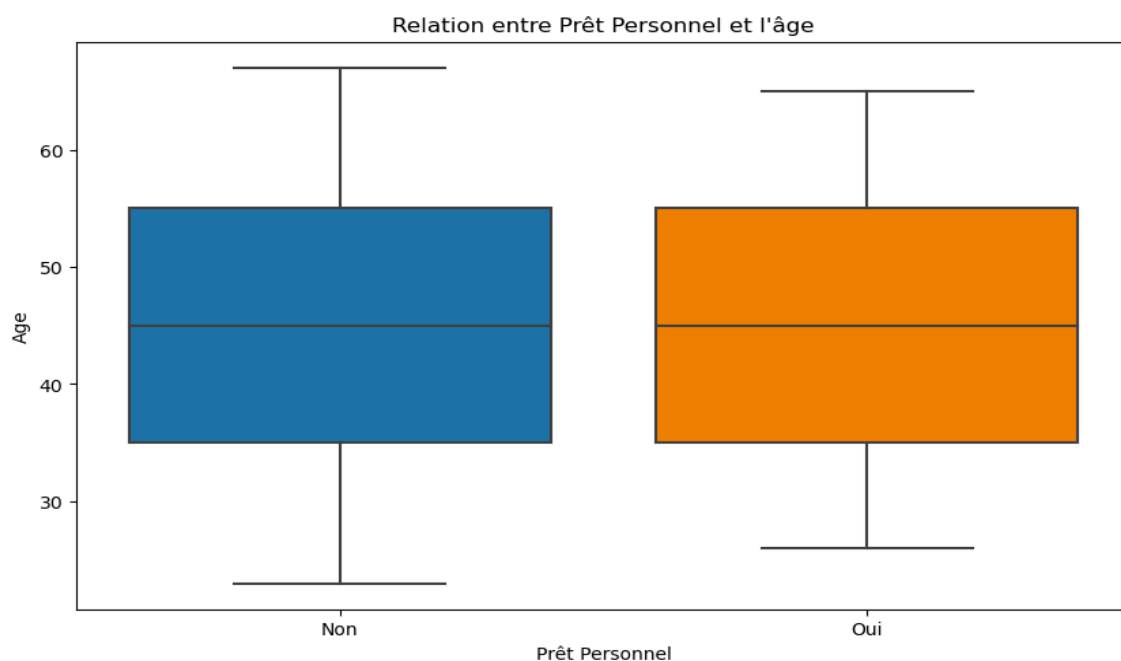


Figure 1 : Boxplot de la relation entre prêt personnel et l'âge

Le boxplot illustre la répartition de l'âge pour les clients avec et sans prêt personnel, montrant des médianes similaires à 45 ans. Les moustaches étendues signalent une dispersion comparable. Malgré une légère différence d'âge moyen chez les souscripteurs, les deux groupes présentent des distributions d'âge relativement similaires. Les tests d'hypothèse confirment l'absence de différences significatives d'âge entre les deux groupes, renforçant l'idée que l'âge seul ne joue pas un rôle majeur dans la décision d'accorder un prêt personnel.

#### ❖ Liaison entre prêt personnel et l'expérience de notre échantillon de clients

L'expérience professionnelle peut jouer un rôle déterminant dans la décision d'accorder un prêt personnel. Une expérience plus étendue pourrait être perçue comme un indicateur de stabilité financière et de capacité de remboursement. Les individus avec une expérience professionnelle substantielle pourraient ainsi présenter une propension plus élevée à se voir accorder un prêt personnel, résultant d'une confiance accrue des prêteurs dans leur capacité à rembourser.

Prêt Personnel	Count	Mean	Std	Min	25%	50%	75%	Max
Non	4415	20.11	11.46	-3	10	20	30	43
Oui	479	19.87	11.58	0	9	20	30	41

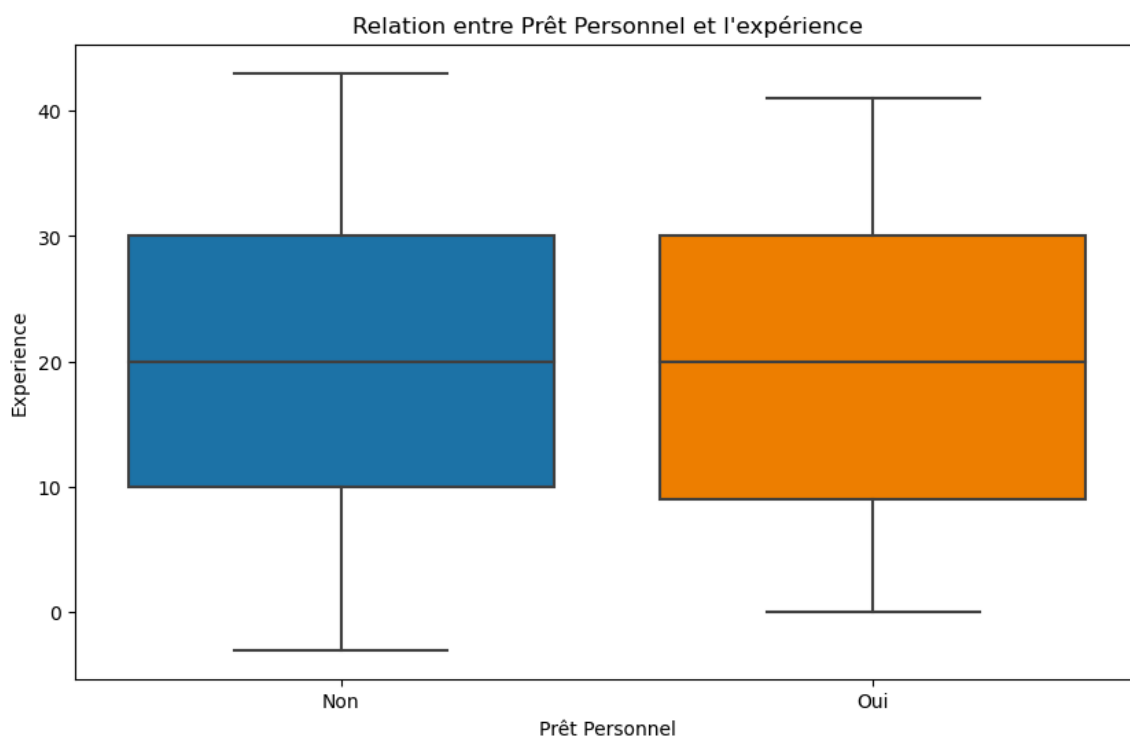


Figure 2 : Boxplot de la relation entre prêt personnel et l'expérience

L'analyse statistique de l'expérience en années par rapport à l'accord de prêt personnel met en lumière une similitude marquée entre les deux groupes. Le boxplot révèle des médianes proches de 20 ans d'expérience, avec des données bien réparties. Malgré une légère différence dans la moyenne, les tests d'hypothèse confirment l'absence de divergences significatives. En somme, cette analyse indique que l'expérience professionnelle, mesurée en années, ne semble pas être un facteur déterminant dans la décision d'octroyer un prêt personnel.

#### ❖ Liaison entre prêt personnel et Taille de la famille de notre échantillon de clients

Intuitivement, la taille de la famille peut être un indicateur de la charge financière globale, influençant potentiellement la capacité à rembourser un prêt. Un nombre élevé de membres dans une famille peut être associé à des dépenses plus importantes, ce qui peut créer des difficultés potentielles pour rembourser un prêt, impactant ainsi la décision d'accorder ou non le prêt personnel.

Prêt Personnel	Count	Mean	Std	Min	25%	50%	75%	Max
Non	4415	2.38	1.15	1	1	2	3	4
Oui	479	2.62	1.11	1	2	3	4	4

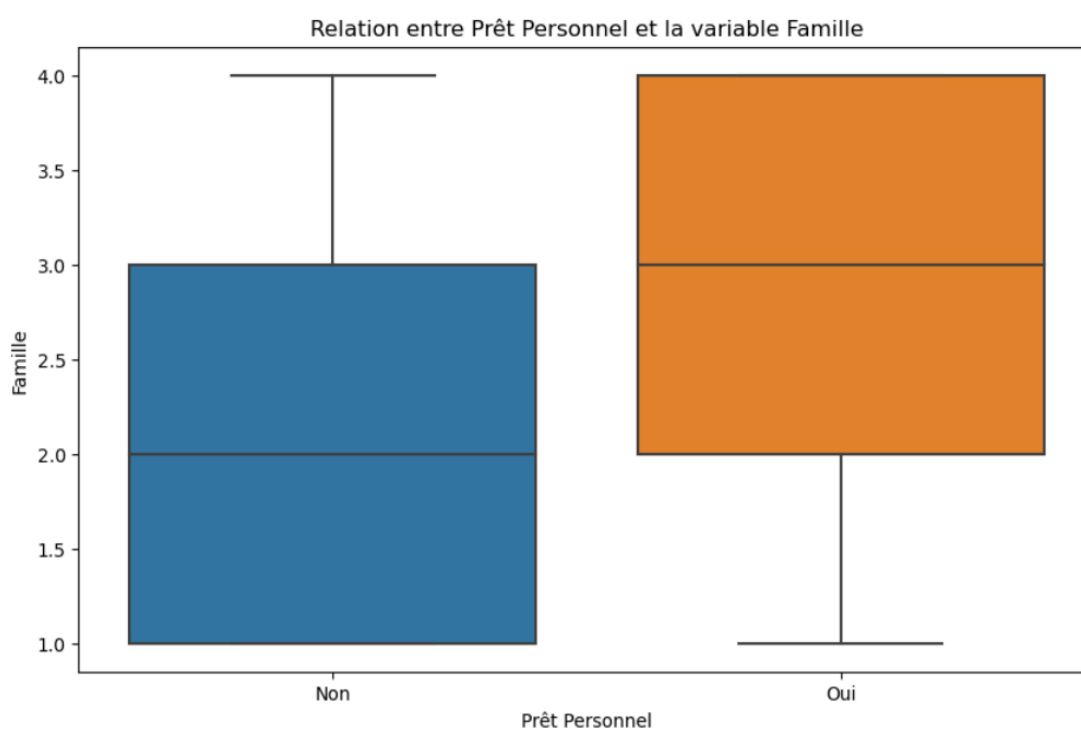


Figure 3 : Boxplot de la relation entre le prêt personnel et le nombre d'individus par famille

Le boxplot révèle des tendances en fonction de la taille de la famille entre les détenteurs et les non-détenteurs de prêt. Les deux groupes présentent des médianes différentes (2 pour les non-détenteurs et 3 pour les détenteurs), une dispersion plus importante est observée chez les détenteurs. Cette variabilité accrue est confirmée par le test statistique, suggérant que la taille de la famille peut influencer la décision d'accorder ou non un prêt personnel.

#### ❖ Liaison entre prêt personnel et l'Éducation de notre échantillon de clients

De manière plausible, le niveau d'éducation peut être un facteur influent dans la décision d'accorder ou non un prêt personnel. Les individus avec un niveau d'éducation plus élevé



pourraient être perçus comme ayant des perspectives de revenus plus stables et des compétences financières accrues, ce qui pourrait augmenter la probabilité d'obtenir un prêt. À l'inverse, ceux avec un niveau d'éducation plus bas pourraient être considérés comme présentant des risques plus élevés, ce qui pourrait influencer négativement la décision d'accorder un prêt.

Éducation/ Prêt	1	2	3
Non	43.99	27.25	28.77
Oui	19.42	37.79	42.80

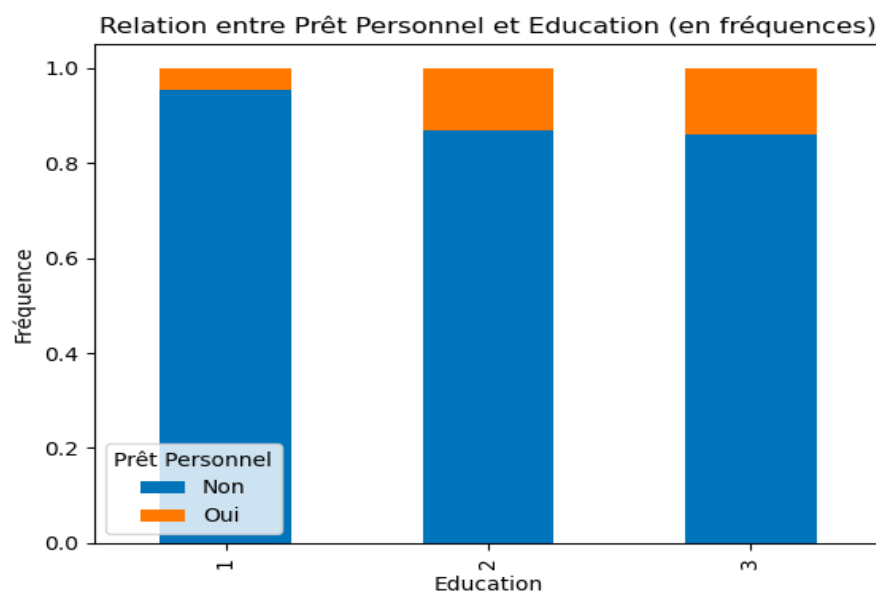


Figure 4: Diagramme en barres de la relation entre le prêt personnel et l'éducation

L'analyse du graphique en barres empilées confirme visuellement que le niveau d'éducation peut influencer la décision d'accorder un prêt personnel, avec une proportion croissante de personnes bénéficiant d'un prêt à mesure que le niveau d'éducation augmente. Cette relation semble significative statistiquement, suggérant que le niveau d'éducation peut jouer un rôle important dans la propension à obtenir un prêt personnel.

## 1-2 Profil financier de nos clients :

### ❖ Liaison entre prêt personnel et le revenu de notre échantillon de clients

Le revenu peut influencer la décision d'avoir un prêt personnel, car un revenu plus élevé offre une capacité financière accrue pour rembourser le prêt. Les clients avec des revenus plus élevés ont généralement une marge de manœuvre financière plus importante

Prêt Personnel	Count	Mean	Std	Min	25%	50%	75%	Max
Non	4415	65.99	40.46	8	35	59	84	224
Oui	479	144.65	31.54	60	122	142	172	203

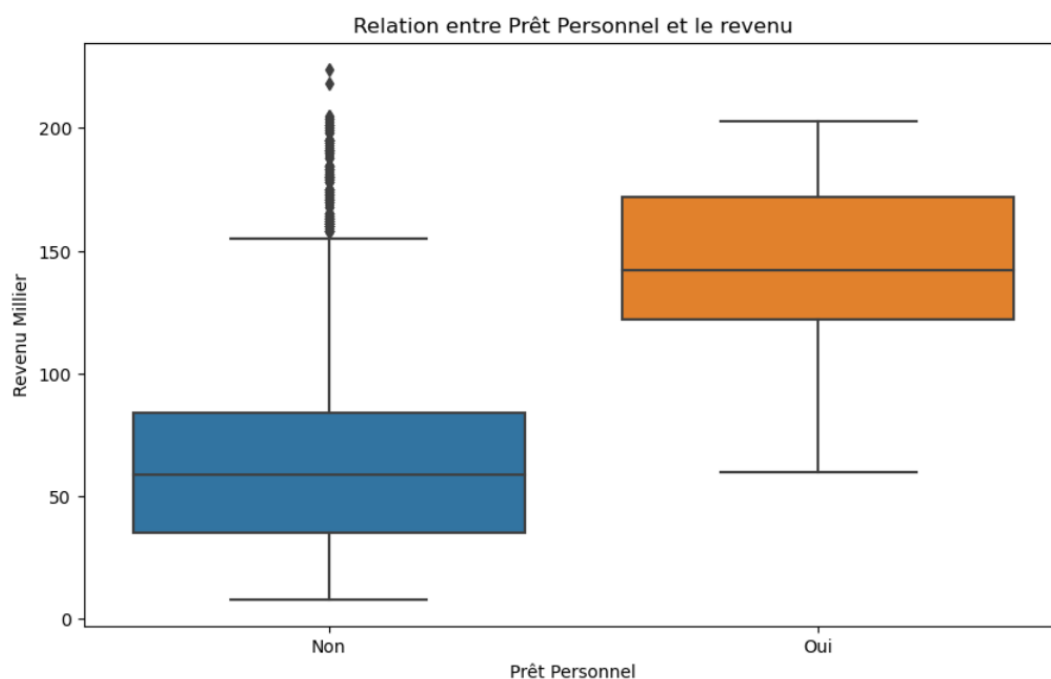


Figure 5: Boxplot de la relation entre prêt personnel et le revenu

Le boxplot souligne une distinction notable dans la distribution du revenu entre les détenteurs et les non-détenteurs de prêt, marquée par une médiane plus élevée chez les premiers (142 milliers d'euros) par rapport aux seconds (59 milliers d'euros). Cette disparité est étayée par une différence significative dans les moyennes de revenu, confirmée par le test statistique

approprié. Ainsi, le revenu émerge comme un facteur clé influençant la décision d’obtenir un prêt personnel au sein de notre clientèle.

❖ **Liaison entre prêt personnel et les dépenses mensuelles de notre échantillon de clients**

Les dépenses mensuelles peuvent influencer la décision d'avoir un prêt personnel en reflétant la capacité financière d'un individu à rembourser le prêt, les montants plus élevés indiquant une marge de remboursement potentiellement plus large.

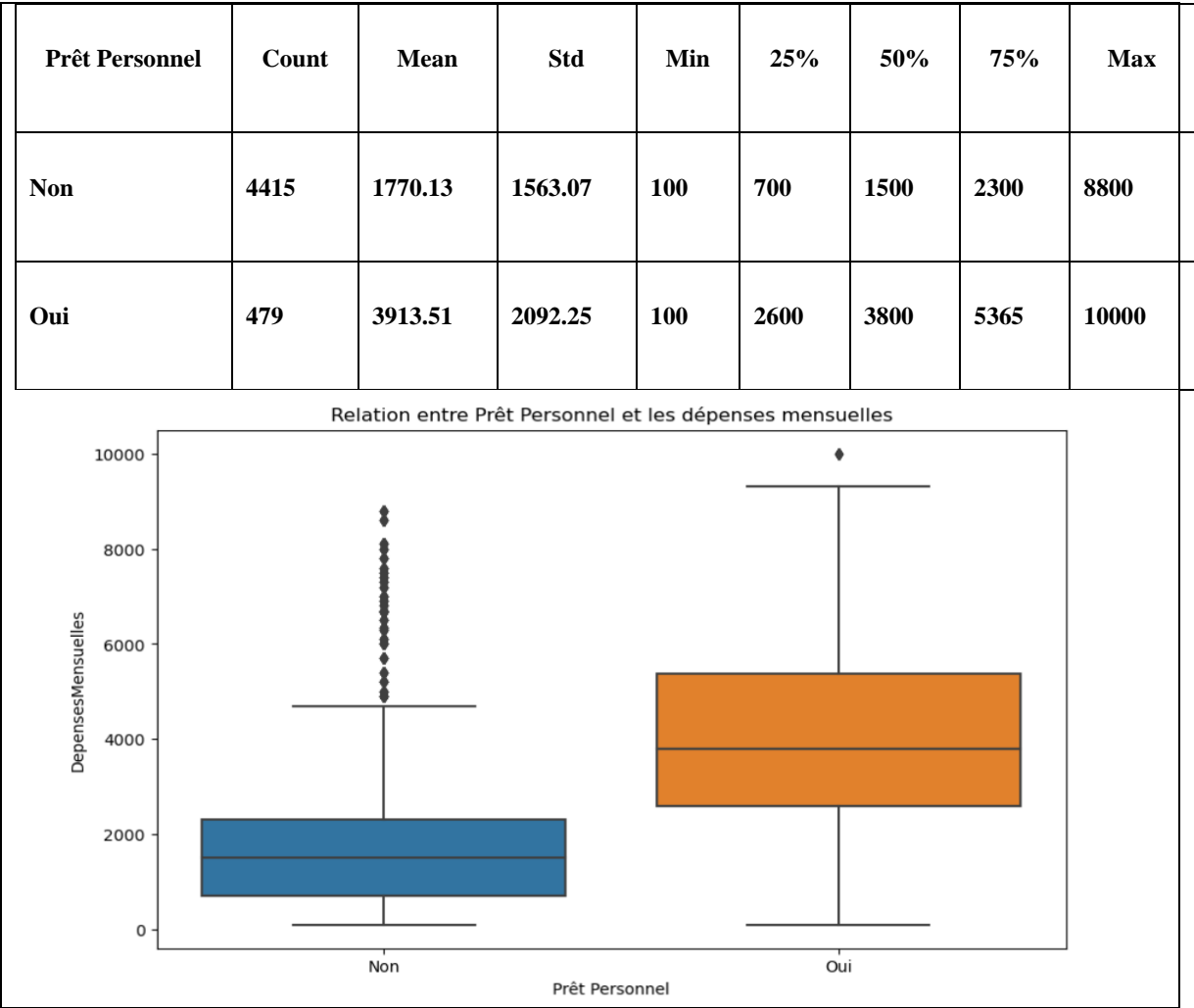


Figure 6 : Boxplot de la relation entre le prêt personnel et les dépenses mensuelles en euro

Le boxplot révèle une distinction significative dans la distribution des dépenses mensuelles entre les détenteurs et les non-détenteurs de prêt. Les détenteurs affichent des dépenses mensuelles plus élevées, avec une médiane de 3800 €, comparée à celle des non-détenteurs, qui est de 1500 €. Cette différence significative est également soutenue par un test statistique,

confirmant que les niveaux de dépenses mensuelles jouent un rôle dans la décision d'accorder un prêt personnel. Ainsi, il semble que les clients avec des dépenses mensuelles plus élevées soient davantage enclins à avoir un prêt personnel.

#### ❖ **Liaison entre prêt personnel et la détention d'un compte épargne logement de notre échantillon de clients**

Intuitivement, le fait de détenir un compte épargne logement peut influencer positivement la décision d'obtenir un prêt personnel. Cela peut être dû au fait que la possession d'un compte épargne logement dénote une certaine stabilité financière et une capacité à épargner.

Prêt Personnel \ EpargneLogement	Non	Oui
Non	96.38%	3.62%
Oui	70.77%	29.23%

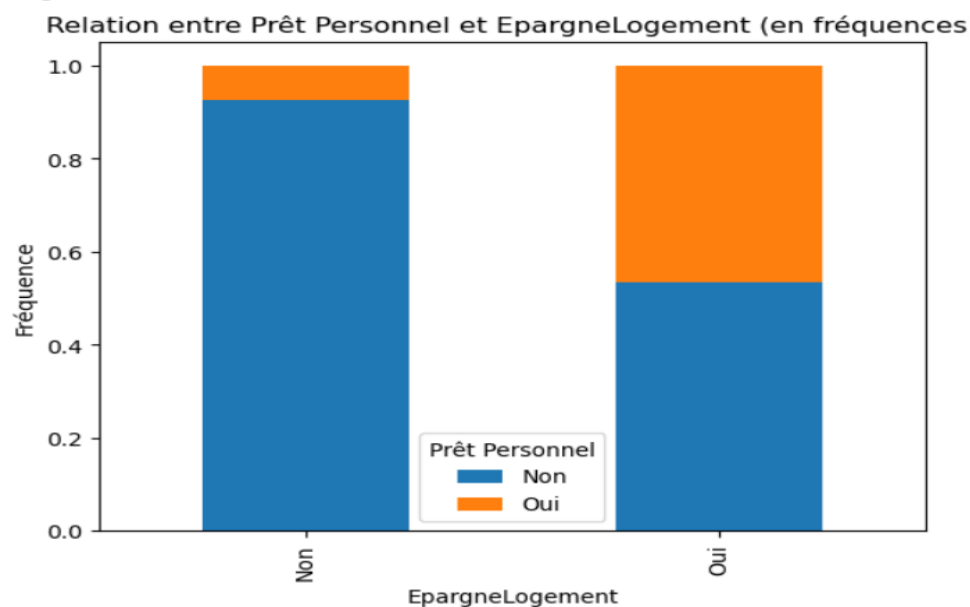


Figure 7: Diagramme en barres de la relation entre prêt personnel et épargne logement

La figure ci-dessus met en évidence une tendance où les individus possédant un compte épargne logement ont une probabilité plus élevée d'obtenir un prêt personnel par rapport à ceux qui n'en ont pas. Cette disparité significative est corroborée par une validation statistique, suggérant que

la détention d'un compte épargne logement est associée à une propension accrue à souscrire à un prêt personnel.

## Partie 2 : Évaluation Préliminaire de l'Importance des Variables

Avant d'entamer la phase de classification, nous avons considéré essentiel d'évaluer l'importance des variables afin d'identifier celles ayant un impact significatif sur notre variable cible, à savoir l'octroi ou non d'un prêt.

Après avoir exploré différents modèles tels que la régression logistique, le modèle probit et l'arbre de décision, notre choix s'est arrêté sur deux modèles. La raison en est que chacun d'eux offre une perspective précise sur l'importance des variables : le modèle probit fournit des coefficients pour chaque variable, tandis que l'arbre de décision offre une représentation visuelle, renforcée par le rapport de cotes qui s'avère également instructif. Ainsi, nous avons opté pour cette approche double pour obtenir une compréhension approfondie de l'impact des variables sur la décision d'accorder un prêt.

### 2-1 Modèle Probit :

Variable	Coefficient	Significativité ( $P >  z $ )
const	-5.3471	0.000
RM	0.0310	0.000
FAM	0.2909	0.000
DM	7.989e-05	0.000
EL_Oui	1.3316	0.000
EDU_2	1.4442	0.000
EDU_3	1.4900	0.000

Le

modèle probit que nous avons ajusté pour prédire la probabilité d'obtention d'un prêt présente des résultats significatifs. Voici une interprétation des principaux éléments du modèle :

- Constante (const): La constante est estimée à -5.35. Cela signifie que lorsque toutes les autres variables sont maintenues constantes, la probabilité d'obtenir un prêt est réduite de 5.35 unités.
- RM (Coefficient : 0.0310): Une augmentation d'une unité de revenu est associée à une augmentation de 0.0310 de probabilité d'obtenir un prêt.
- FAM (coefficient : 0.2909): La variable famille affiche un coefficient de 0.2909. Cela suggère qu'une augmentation d'une unité dans la situation familiale, qui n'est autre que le nombre d'individus, est associée à une augmentation de 0.2909 dans la probabilité d'obtenir un prêt
- DM (Coefficient : 7.989e-05): Une unité supplémentaire des dépenses mensuels est associée à une augmentation de 7.989e-05 dans la probabilité d'obtenir un prêt.
- EL\_Oui (Coefficient : 1.3316): Pour la variable épargne logement, la présence de ce dernier est associée à une augmentation de 1.3316 d'avoir un prêt.
- EDU\_2 (Coefficient : 1.4442) et EDU\_3 (Coefficient : 1.4900): le résultat de la variable niveau d'éducation est significatif; un niveau d'éducation moyen et élevé est associé à une augmentation respective de 1.4442 et 1.4900 dans la probabilité d'obtenir un prêt. Ainsi, avoir un niveau d'éducation plus élevé semble être un facteur positif contribuant à une plus grande probabilité d'accord de prêt.

Le R2 est de 0.8226 indique la proportion de la variance expliquée par le modèle, montrant une bonne adéquation aux données. Toutes les variables présentent une significativité statistique (P-value < 0.05), soulignant leur importance dans la prédiction du résultat.

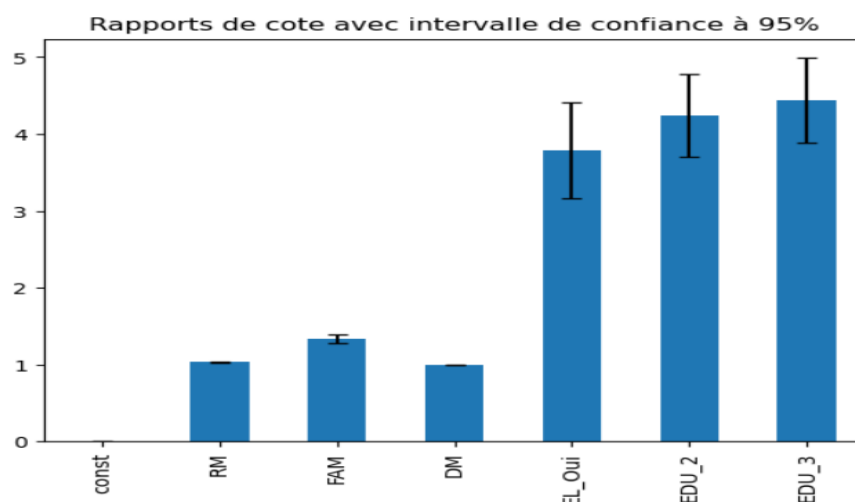


Figure 8 : Diagramme en barres des rapports de cote

Dans le contexte du modèle probit, l'analyse des rapports de cotes met en lumière le rôle prédominant de l'éducation dans la détermination des chances d'obtenir un prêt. En tête de liste, les niveaux d'éducation moyen et élevé présentent des rapports de cotes significativement élevés, indiquant que les individus ayant ces niveaux d'éducation ont une probabilité substantiellement plus grande d'obtenir un prêt par rapport à ceux ayant un niveau d'éducation plus bas.

Ensuite, le compte d'épargne de logement émerge comme un facteur important, bien que sa contribution soit moins prononcée que celle de l'éducation. Les individus ayant un compte d'épargne de logement bénéficient d'une probabilité accrue d'obtenir un prêt.

Le nombre de familles joue également un rôle significatif, bien qu'avec une influence moindre par rapport aux variables précédentes. Une augmentation du nombre de familles est associée à une augmentation notable des chances d'accord de prêt.

Enfin, le revenu mensuel et les dépenses présentent des rapports de cotes positifs, leur impact sur l'obtention de prêts est moins prononcé par rapport aux autres variables.

## 2-2 Arbre de décision :

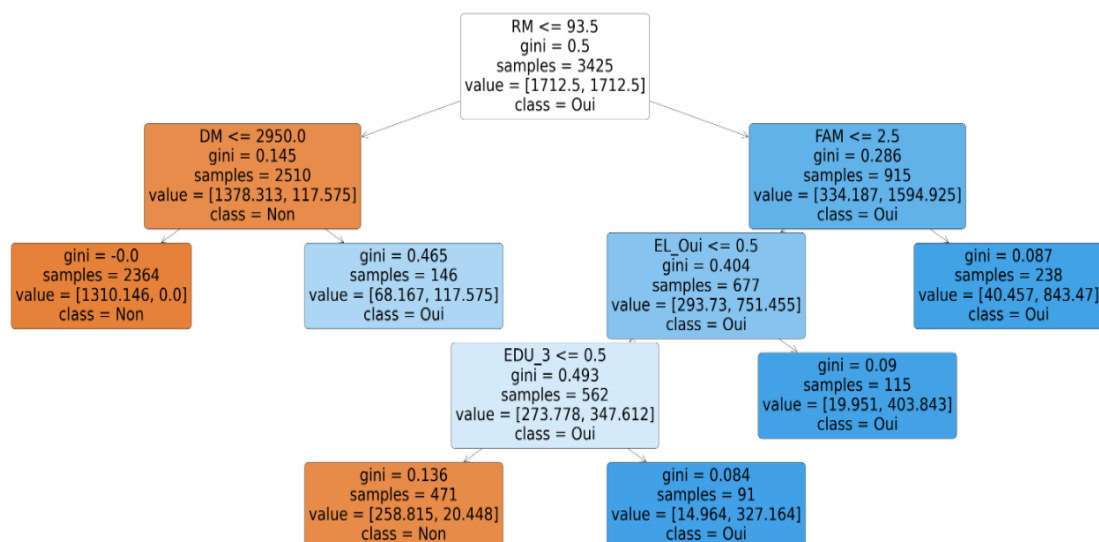


Figure 9 : Arbre de décision

Pour interpréter cet arbre de décision, nous pouvons envisager deux scénarios possibles en fonction des différentes conditions établies par le modèle.

En combinant ces deux scénarios, le modèle offre une approche séquentielle pour évaluer la probabilité d'obtenir un prêt en fonction du revenu, du nombre d'individus par famille, de la détention d'un compte de logement, du niveau d'éducation, et des dépenses mensuelles.

Le modèle hiérarchique prend comme critère initial le revenu. Deux scénarios distincts se dégagent à ce stade.

Dans le premier scénario, si le revenu est inférieur ou égal à 93.5 mille euros, le modèle examine ensuite le nombre d'individus par famille. Si ce nombre est inférieur ou égal à 2.5, l'individu a une probabilité d'obtenir un prêt. En revanche, si le nombre d'individus par famille est supérieur à 2.5, le modèle examine la détention d'un compte de logement. Si le compte de logement est détenu, l'individu a une probabilité d'obtenir un prêt. Dans le cas contraire, le modèle vérifie le niveau d'études. Si le niveau d'études est élevé, l'individu a une probabilité d'obtenir un prêt, sinon ses chances sont réduites.

Dans le deuxième scénario, si le revenu est supérieur à 93.5 mille euros, le modèle analyse les dépenses mensuelles. Si les dépenses mensuelles sont inférieures ou égales à 2950 euros, l'individu a une probabilité plus élevée d'obtenir un prêt. Cependant, si les dépenses mensuelles dépassent 2950 euros, ses chances d'obtenir un prêt sont réduites. Ces scénarios représentent des étapes séquentielles dans le processus de prise de décision du modèle en matière d'octroi de prêts, mettant en évidence les conditions spécifiques à considérer à chaque étape.



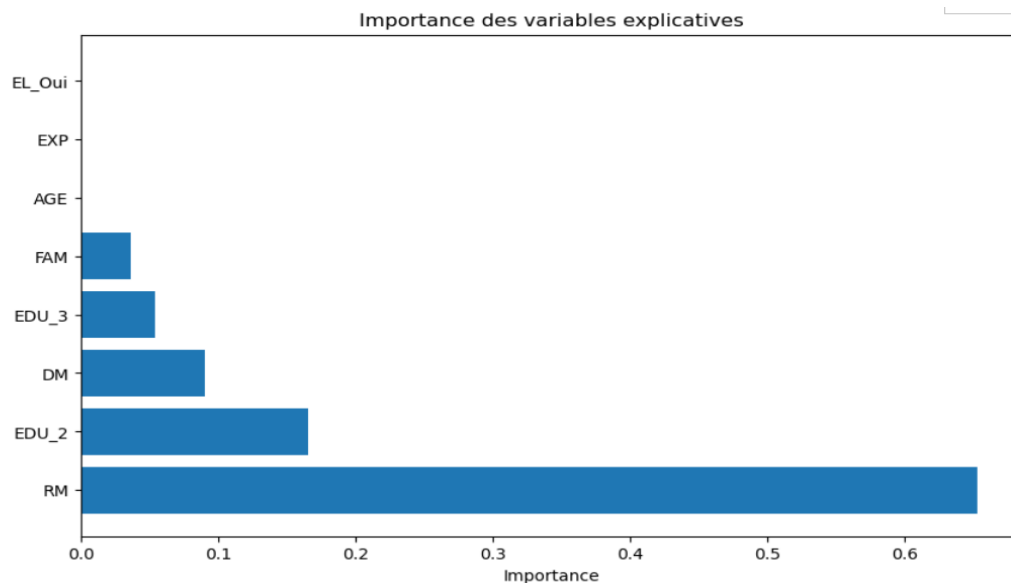


Figure 10 : Diagramme des barres qui représente l'importance des variables explicatives

La variable la plus cruciale dans la décision d'obtenir un prêt est le revenu, selon le modèle, suivi par le niveau d'éducation et les dépenses mensuelles, dans un ordre décroissant d'importance. Cela suggère que, parmi les variables examinées, le revenu a le plus grand impact sur les chances d'obtenir un prêt, suivi de près par le niveau d'éducation et les dépenses mensuelles, qui restent également des facteurs significatifs dans le processus de décision lié aux prêts.

Les variables AGE, EXP, n'ont pas d'impact significatif sur les chances d'obtenir un prêt selon ce modèle.

### Partie 3 : MISE EN COMPETITION DES MODELES ET VALIDATION

L'évaluation des modèles, avant et après agrégation, se concentre principalement sur l'AUC et le rappel. Cette approche mesure l'impact de l'agrégation sur les performances globales sans entrer dans les détails de chaque mesure. L'AUC évalue la capacité de discrimination entre les classes, tandis que le rappel est examiné de près, en particulier dans des situations de déséquilibre de classe. Ces analyses simplifiées guident la prise de décision pour choisir le modèle final en se basant sur des indicateurs clés de performance.

	Sans Agrégation				Avec Agrégation (Bagging)			
Modèle	Exactitude	Précision	Rappel	AUC	Exactitude	Précision	Rappel	AUC
Logit	0.9	0.49	0.92	0.9657	0.9	0.49	0.92	0.90
Probit	0.9	0.49	0.92	0.9658	0.96	0.87	0.64	0.8142
Cloglog	0.91	0.53	0.88	0.9604	-	-	-	-
Arbre de décision	0.942	0.631	0.98	0.9786	0.80	0.337	1.0	0.9795
SVM	0.844	0.379	0.9227	0.9480				

L'AUC, ou aire sous la courbe ROC (Receiver Operating Characteristic), est une mesure de la performance d'un modèle de classification, évaluant sa capacité à discriminer entre les classes. La courbe ROC représente la relation entre la sensibilité (taux de vrais positifs) et la spécificité (taux de faux positifs) du modèle pour différents seuils de classification. Une AUC plus élevée indique une meilleure capacité du modèle à distinguer entre les classes.

#### ❖ Comparaison de l'AUC des Modèles :

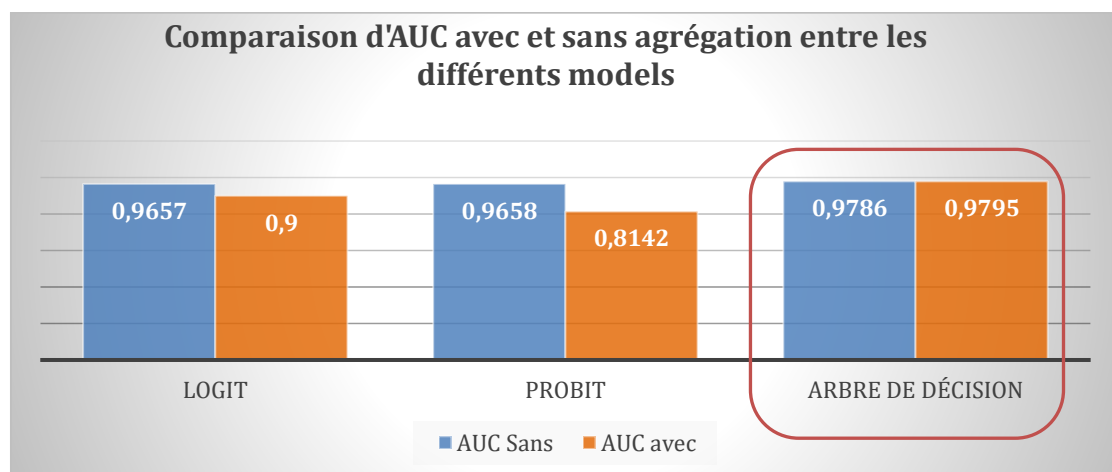


Figure 12 : Diagramme en barres qui représente une comparaison d'AUC avec et sans agrégation entre les différents modèles

- Avant agrégation, l'Arbre de décision a affiché le meilleur AUC à 0.9786.
- Avec agrégation, l'Arbre de décision maintient un AUC élevé à 0.9795, tandis que les autres modèles montrent une variation dans leurs performances, notamment une diminution pour Logit et Probit.

### Choix du Modèle en se basant sur l'AUC :

- En termes d'AUC, l'arbre de décision avec agrégation semble avoir la meilleure performance discriminatoire avec une AUC de 0.9795.

### ❖ Comparaison du Rappel des Modèles :

Étant donné le déséquilibre de classe dans notre ensemble de données, nous avons pris en compte cette considération en utilisant des modèles pondérés, où des poids ont été ajoutés pour équilibrer l'impact des classes. Cette approche permet au modèle de mieux tenir compte des exemples de la classe minoritaire, améliorant ainsi sa performance globale.

En raison de cette pondération et du déséquilibre observé, l'importance du rappel (sensibilité) devient cruciale. **Le rappel mesure la capacité du modèle à bien identifier les cas positifs, particulièrement essentielle lorsque la classe représentant ceux qui ont obtenu un prêt est sous-représentée. Cependant, il est crucial de ne pas perdre de vue l'objectif principal, qui est souvent de bien prédire les cas positifs, c'est-à-dire ceux qui ont obtenu un prêt.**

Il est donc impératif de ne pas se limiter à l'AUC lors de l'évaluation des performances. En comparant les modèles, privilégiez celui qui offre un bon équilibre entre l'AUC et le rappel, assurant ainsi une meilleure détection des cas importants malgré le déséquilibre de classe.

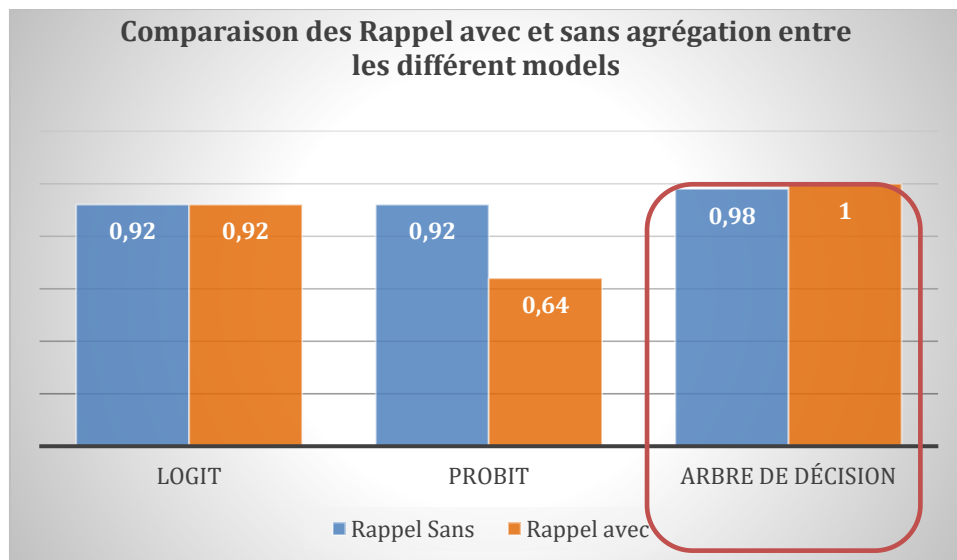


Figure 13 : Diagramme en barres qui représente une comparaison du Rappel avec et sans agrégation entre les différents modèles

- Sans agrégation, l'arbre de décision présente le rappel le plus élevé (0.98), indiquant une excellente capacité à détecter les vrais positifs.
- Avec agrégation, l'arbre de décision maintient un rappel élevé (1.0), tandis que d'autres modèles, tels que Probit, voient une diminution significative du rappel.

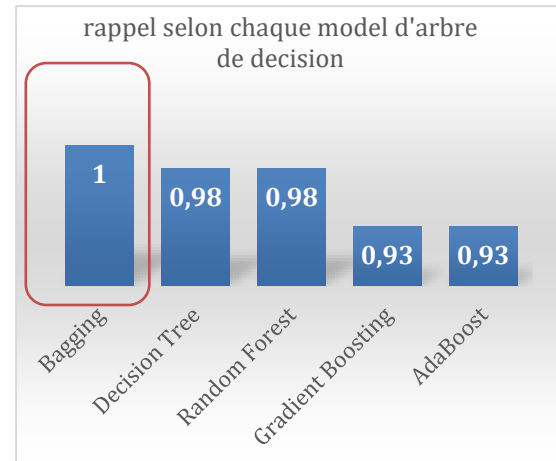
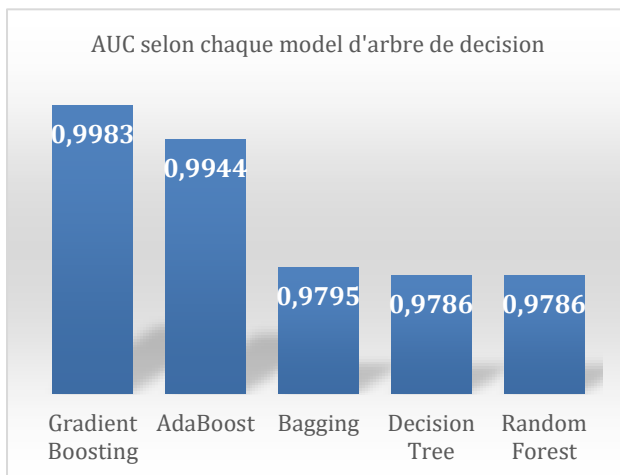
#### Choix du Modèle en se basant sur le Rappel :

- En considérant l'importance du rappel, l'arbre de décision, en particulier avec agrégation, semble offrir la meilleure performance pour détecter les cas positifs, même en tenant compte du déséquilibre de classe.

L'arbre de décision avec agrégation a un rappel de 1.0, cela suggère que ce modèle a correctement identifié tous les individus qui ont obtenu un prêt.

Dans notre exploration de méthodes d'agrégation pour les arbres de décision, nous avons évalué plusieurs techniques, dont le Bagging, Random Forest, AdaBoost, et Gradient Boosting. L'objectif était de déterminer si ces méthodes d'agrégation pouvaient améliorer les performances par rapport à un simple arbre de décision.

Comparaison des modèles arbres avec agrégation				
Modèle	Exactitude	Précision	Rappel	AUC
<b>Decision Tree</b>	<b>0.942</b>	<b>0.631</b>	<b>0.98</b>	<b>0.9786</b>
<b>Bagging</b>	<b>0.807</b>	<b>0.337</b>	<b>1.0</b>	<b>0.9795</b>
<b>Random Forest</b>	<b>0.94</b>	<b>0.63</b>	<b>0.98</b>	<b>0.9786</b>
<b>AdaBoost</b>	<b>0.99</b>	<b>0.97</b>	<b>0.93</b>	<b>0.9944</b>
<b>Gradient Boosting</b>	<b>0.98</b>	<b>0.94</b>	<b>0.93</b>	<b>0.9983</b>



En comparant les modèles d'arbres de décision avec agrégation, bien que les méthodes de Gradient Boosting et AdaBoosting aient présenté des scores élevés par rapport au Bagging et au Random Forest, il est crucial de noter que, du point de vue du rappel, le Bagging a atteint une performance parfaite avec un score de 1 pour l'identification de tous les cas positifs. Cela met en évidence l'efficacité exceptionnelle du Bagging dans la détection exhaustive des cas positifs, malgré la possibilité que d'autres méthodes d'agrégation aient affiché des scores globaux plus élevés.

En se basant sur l'importance du rappel, le Bagging de l'arbre de décision restera toujours notre choix privilégié, car il parvient à identifier tous les cas positifs.

## Partie 4 : Prédiction de l'octroi du prêt avec les nouvelles données

En utilisant le modèle sélectionné, à savoir l'arbre avec agrégation (Bagging), nous avons effectué des prédictions pour de nouveaux clients afin d'évaluer sa robustesse dans la prédiction de la probabilité d'obtention d'un prêt. Cette étape vise à tester la capacité du modèle à généraliser ses prédictions à des données inédites, permettant ainsi une évaluation de sa performance hors de l'échantillon d'entraînement initial.

Total des clients auxquels un prêt sera accordé	9
Total des clients auxquels un prêt sera refusé	21

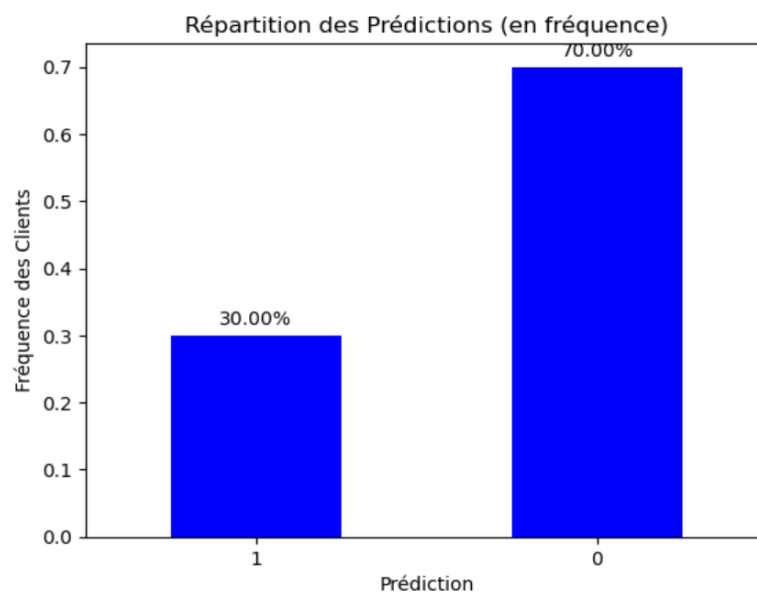


Figure 14 : Diagramme en barres représentant la prédiction de l'octroi du prêt

En analysant les prédictions effectuées avec le modèle, on peut observer que les individus nouvellement évalués ont été catégorisés en deux classes : ceux prévus pour obtenir un prêt (1) et ceux ne l'obtenant pas (0).

En se basant sur le rapport de côtes du modèle (voir Figure 10), on constate que les variables "revenu" et "niveau d'éducation" semblent être les plus déterminantes pour la décision d'accorder un prêt ou non. Ces deux variables exercent une influence significative sur la probabilité d'obtention d'un prêt, indiquant que les individus avec un revenu élevé et un niveau d'éducation élevé sont plus susceptibles de recevoir un prêt.

Par ailleurs, les variables "dépenses mensuelles" et "nombre de membres de la famille" semblent également jouer un rôle dans la décision, mais leur impact est moins prononcé comparé aux deux premières variables.

En examinant spécifiquement les nouveaux individus qui ont obtenu un prêt selon les prédictions du modèle, on confirme que leur profil se caractérise effectivement par un revenu élevé et un niveau d'étude élevé. Ces résultats renforcent la validité du modèle en identifiant les caractéristiques clés associées à l'obtention d'un prêt.

## **Conclusion :**

En synthèse, après avoir exploré divers modèles grâce à l'apprentissage automatique, notre choix s'est finalement arrêté sur le modèle d'arbre de décision avec agrégation (Bagging). Ce modèle a démontré une performance exceptionnelle dans la prédiction de nouveaux profils pour déterminer s'ils sont éligibles à un prêt. En utilisant cette approche, nous avons atteint l'objectif initial fixé, permettant une réponse précise à la question cruciale de l'éligibilité au prêt pour de nouveaux clients. Le modèle sélectionné offre non seulement une haute précision, mais également une capacité robuste à généraliser et à s'adapter à de nouvelles données, renforçant ainsi sa fiabilité dans un contexte financier.