



Data science et Risque réglementaire

Mémoire de Master MSIC
soutenu le 5 décembre 2022 par Youssef AIT TAHAR

Directeur de mémoire : Samuel PARFOURU
Membre du JURY : Selmin NURCAN



Contenu de la soutenance

 **Prolongements possibles** **8**

1 **Problématique** 

 **Limites de la recherche** **7**

2 **Objectif et contexte du mémoire** 

 **Solutions proposées** **6**

3 **Etat de l'art** 

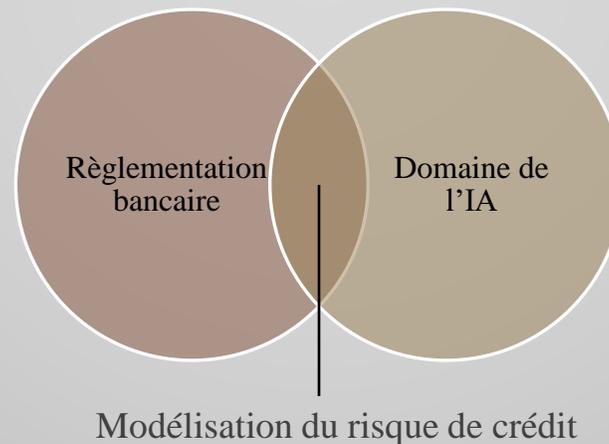
 **Principaux résultats** **5**

4 **Méthodologie de recherche** 



1 Problématique

« Les banques peuvent-elles optimiser leur coût du risque par l'utilisation de Machine Learning dans la modélisation interne du risque de crédit en respectant le cadre réglementaire imposé par le superviseur ? »

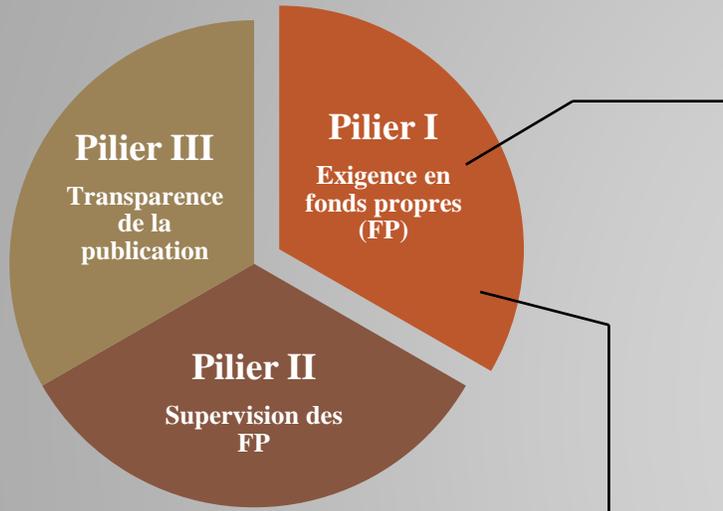


2

Objectif et contexte du mémoire

- Notre problématique est orienté sur les aspects manquants dans la revue de la littérature sur la compatibilité entre les méthodes d'apprentissage automatique dans l'évaluation des risques et l'exigence de la réglementation bancaire.
- Les institutions bancaires avec leur politique d'économie du risque s'intéressent que depuis peu de temps à cette alternative aux méthodes statistiques classique.
- A travers notre recherche, nous allons essayer d'apporter des analyses complémentaires, des débuts de solutions en s'appuyant des discussions avec des professionnels et collaborateurs sur le terrain.

La réglementation bancaire sous Bâle III – Focus sur le Pilier I



Méthodes d'estimation des fonds propres réglementaires en couverture des encours pondérés :

- Risque de crédit : perte financière potentielle en cas de défaut des débiteurs
- Risque de marché : perte financière résultant des fluctuations des prix des instruments financiers
- Risque opérationnel : pertes provenant de processus internes inadéquats ou défaillants, de personnes et systèmes ou d'événements externes

Source : (ACPR Bâle III, 2018)

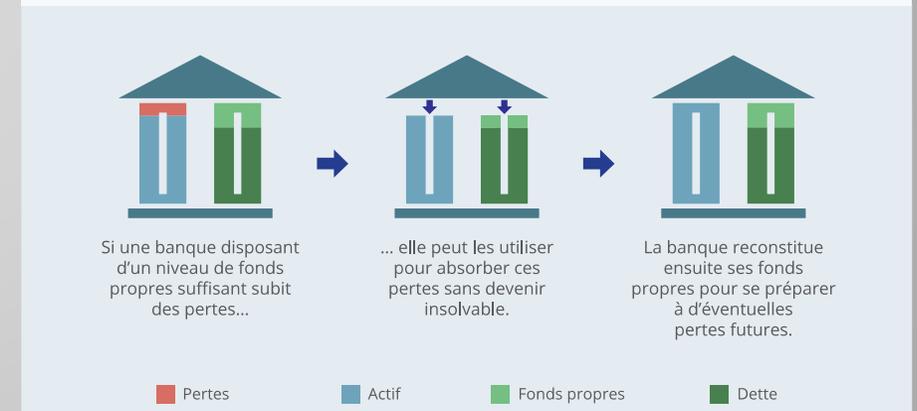
D'un point de vue bilantiel :

Passif : Exigences minimales de fonds propres (EFP) à 8 % de leurs actifs pondérés des risques – C'est le coussin financier

Actif : les actifs de la banque multiplié par leurs différents facteurs de risque (= pondération du risque = RWA). Ces facteurs de risque (PD,LGD,M,EAD) indiquent le degré de risque perçu de certains types d'actifs

Source : (European Banking Authority, 2022)

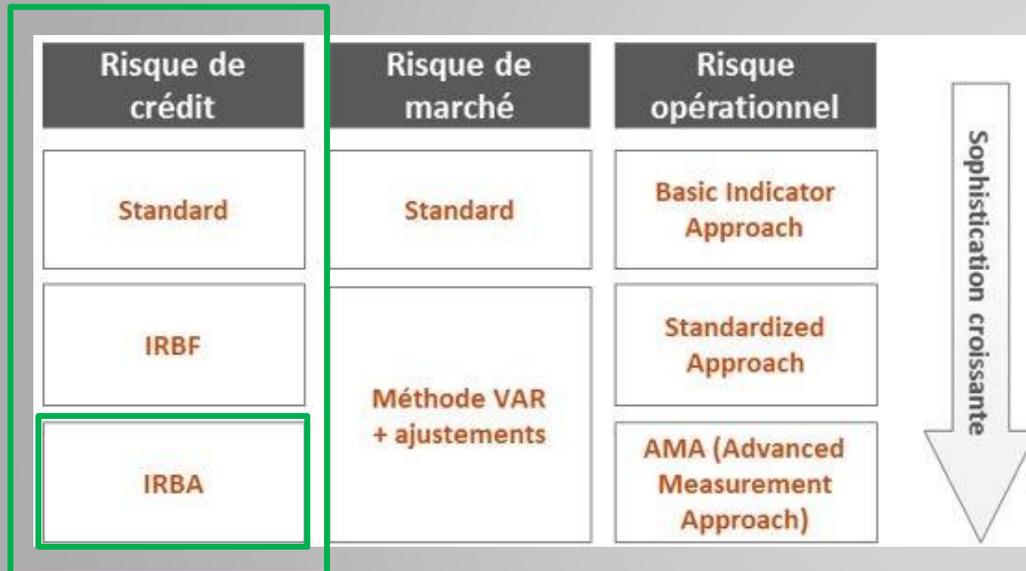
Les banques détiennent des fonds propres pour absorber les pertes



Source : ACPR Normes, 2017)

Méthodes envisageables pour le calcul des risques

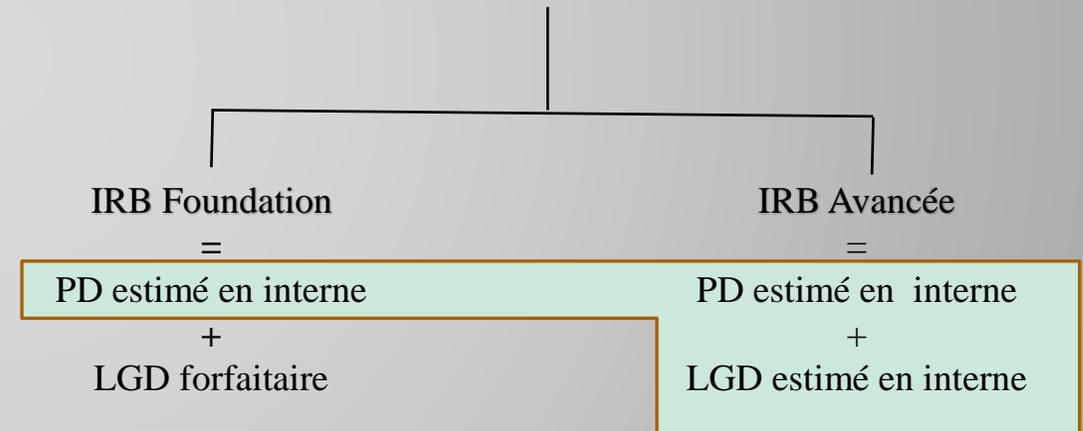
Le risque de crédit constitue aujourd'hui plus de 75% des portefeuilles des banques.



Source : (ACPR Bâle III, 2018)

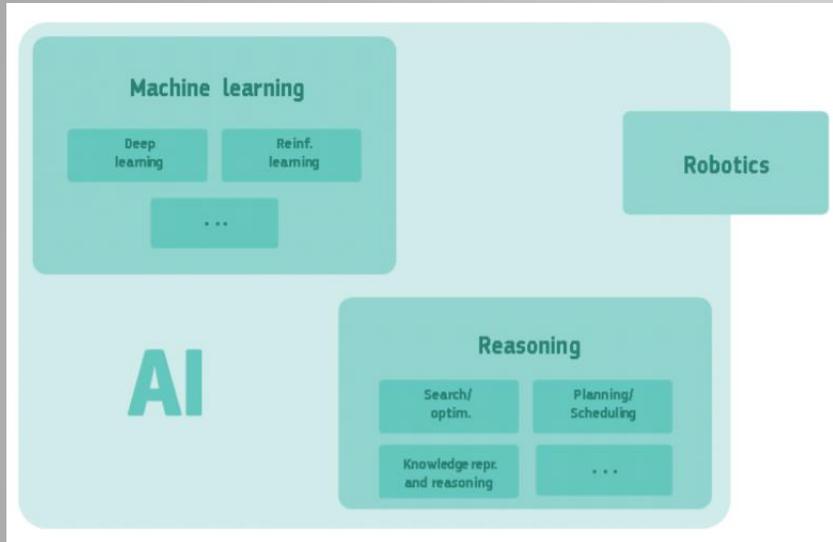
Pourquoi choisir la modélisation interne ?

- Le modèle interne est l'approche la plus précise pour chaque paramètre afin de réduire les exigences de fonds propres tout en conservant la limite du ratio de McDonough de 8%...
- Deux méthodes se distinguent au sein de l'approche sophistiquée



Nouveau terrain pour répondre aux exigences réglementaires à partir de modèles de d'apprentissage automatique (Machine Learning). Les approches sont nouvelles et les estimations fournissent des résultats plus efficaces avec l'algorithme de classification « Gradient Boosting » (Grzybowska and Karwański, 2020).

La conception du Machine Learning



Source : (European Commission, 2020)

- L'IA fait son apparition à la moitié du XXème siècle, désigné par John McCarthy lors d'un conseil national de la recherche aux États-Unis présenté à l'université de Dartmouth en 1956 (Council, 1999)
- L'IA a pour vocation de rendre les machines intelligentes, une qualité qui permet à une entité de fonctionner correctement et de manière prévisible dans son environnement (Nilsson, 2010)
- Cinq axes sont évoqués pour la définir : structure, comportement, capacité, fonction et par principe, Tous par ordre croissant de généralité et décroissant de spécificité (Wang, 2008).

Simuler la pensée humaine

- Test de Turing, L'examineur fait face à deux interlocuteurs, l'un est un ordinateur et l'autre est un humain. En échangeant des textes, il doit savoir lequel des deux est la machine. Si cela échoue, l'ordinateur a réussi le test (Ganascia, 2017)

Les risques et les biais du ML : L'explicabilité

- L'explicabilité mène à une analyse sémantique complète qui consiste à comprendre des données pour en extraire des informations et agir avec raisonnement. Il doit être accessible à l'utilisateur, qu'il soit novice ou technique, quel que soit son degré d'expertise (Gilpin et al., 2018).

- l'IA explicable (XIA) ne repose pas seulement dans la justification des décisions de modèle, mais elle vise à éclairer chaque composante du processus de solution. L'intérêt est de rendre les modèles transparents et compréhensibles pour traiter les problèmes de l'injustice (Gonen et al., 2021).

- L'explicabilité est requise lorsqu'un modèle est validé avant d'être mis en œuvre et déployés dans le système de production. Les prises de décisions importantes qu'elles soient stratégiques ou opérationnelles nécessitent une explicabilité (Martens et al., 2009)



Transparence
des modèles



Générer de la
confiance



Respecter la
réglementation



Comprendre
la prise des
décisions



Améliorer la
performance
des modèles



Réduire les
biais éthiques
et moraux

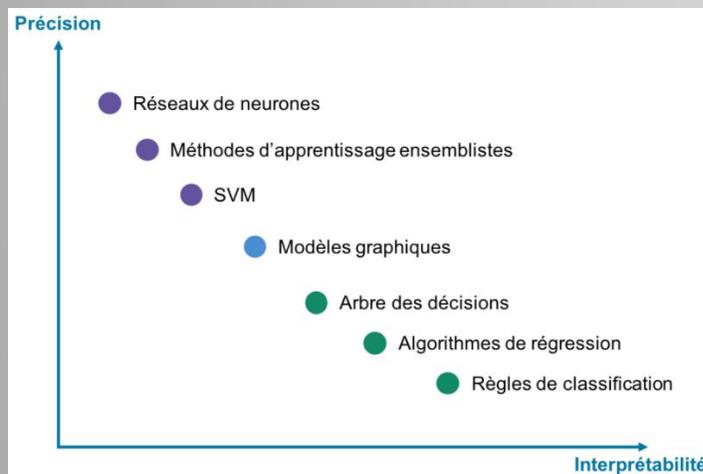
- Les explications fournis aux parties prenantes assure la validité du modèle et ils garantissent la robustesse des processus pour contrôler les modèles dans la prise de décisions (Guidotti et al., 2019).

- Explicabilité = compréhension technique et objective du fonctionnement d'un algorithme (Krishnan, 2020).

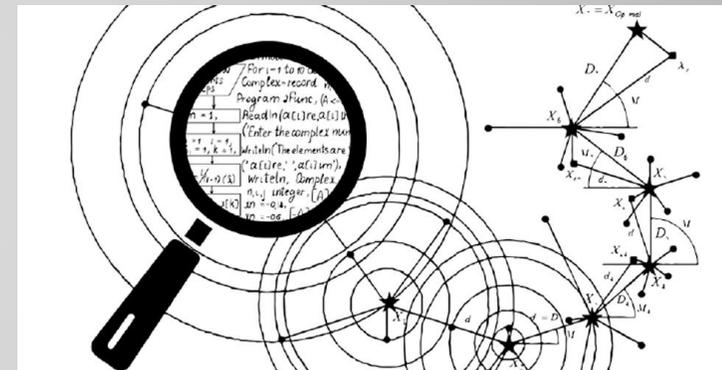
Les risques et les biais du ML : L'interprétabilité et la transparence

- L'interprétabilité consiste à construire le raisonnement algorithmique de l'apprentissage automatique et la structure interne des données dans un registre que les datascientist peuvent interpréter (Gilpin et al., 2018).
- La transparence est associée à la traçabilité par le fait que les données et les différents processus permettent au système de prendre des décisions, y compris le recensement des données et leurs labellisations (Ananny and Crawford, 2018).

Classification de l'interprétabilité des algorithmes



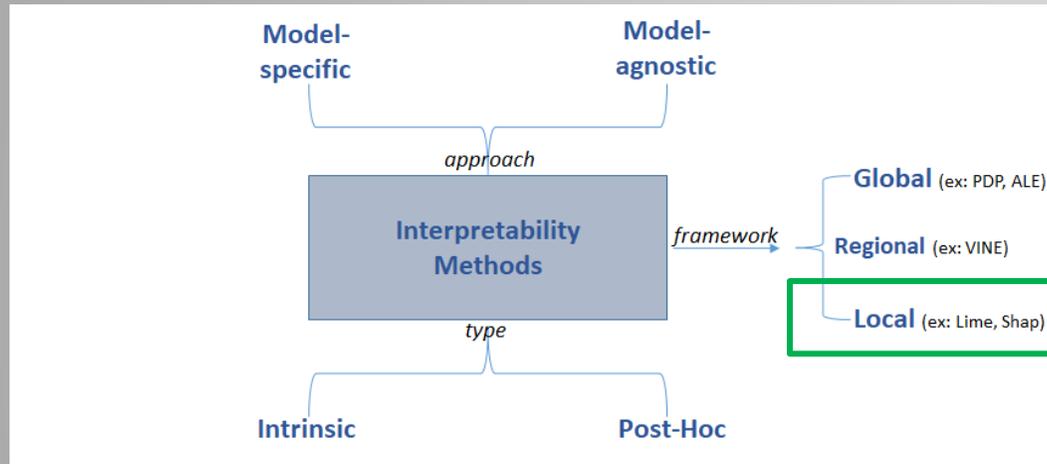
Source : (Dam et al., 2018)



- L'instauration de la transparence est manifeste dans le sens où elle construit nos pensées, et le fait de comprendre les systèmes algorithmiques contribuera à apporter de nouvelles normes d'un point de vue réglementaire dans la gouvernance de l'IA (European Commission, 2020)

Reduction d'opacité des systèmes de ML : Des explications locales indépendantes

Différentes catégories d'interprétabilité des modèles



Source : (Ribeiro, Singh, et Guestrin 2016)

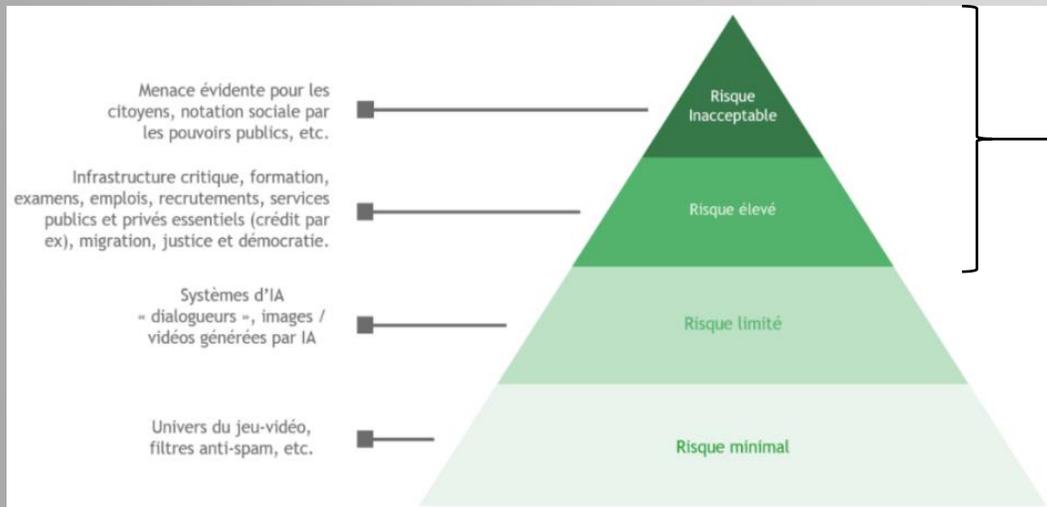
SHAP pourrait ainsi être une méthode d'interprétabilité des modèles répondant aux exigences du "droit à l'explication" instauré par le RGPD.

La méthode SHAP fournit une explication de la prédiction faite par un modèle quelconque en attribuant une valeur de contribution à chaque variable utilisée, contrairement à LIME qui renvoie une réponse plus concise, en pénalisant les modèles complexes <https://arxiv.org/pdf/2007.12919.pdf>

- **Méthode LIME** : Procédure indépendante de chaque modèle qui les opère comme des boîtes noires et les explique localement. On cherche à comprendre le comportement du modèle original (Ribeiro et al., 2016)
 - Le souci rencontré avec la méthode est le fort impact qualitatif quant à la crédibilité du choix approximatif des bornes de décisions locales émises par la boîte noire (Baehrens et al., 2009).
- **Méthode SHAP** : Procédure d'explication additive basée sur la théorie des jeux coopératifs pour distribuer une valeur dit « Shapley », ce résultat final est obtenu selon chaque critère évaluée durant le processus. Les valeurs de Shapley confortent l'intelligibilité des modèles comme les algorithmes de régressions linéaires (Lundberg and Lee, 2017).

Un cadre réglementaire et éthique s'impose

- Proposition sur l'approche de la réglementation basée sur les risques en distinguant les différentes formes de l'utilisation des systèmes l'IA pour minimiser les risques jusqu'à tolérer un niveau de risque faibles (European Commission, 2020).
- la confiance des systèmes est primordiale n'est pas un luxe mais une obligation donner de l'intelligibilité aux processus et aux résultats (European Commission, 2020).

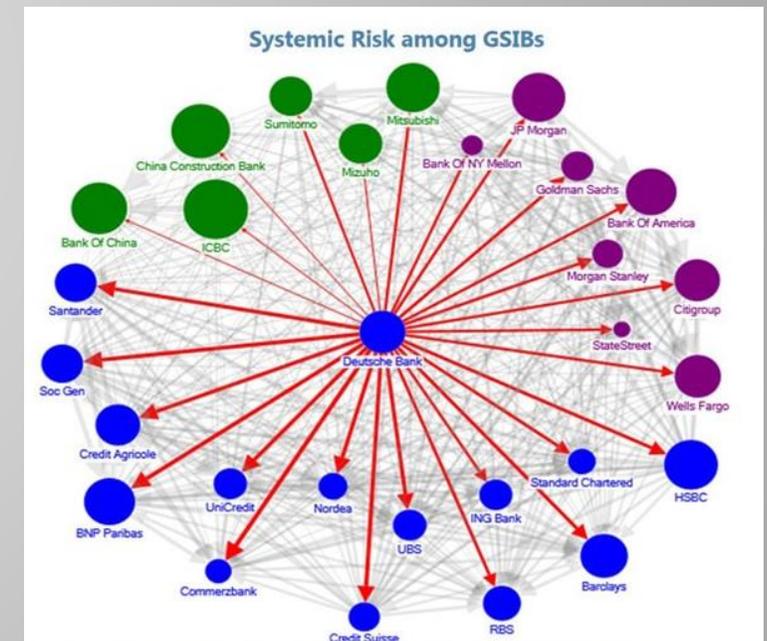


Source : Source : (Studer et al., 2021)

Risque systémique ou effet domino si le SIA échappe aux contrôles d'une banque sur l'aspect des EFP

L'audit interne, le garant de ces systèmes en matière de gouvernance, de contrôle et de risque

- Loin de là, nombre des questions soulevées par les algorithmes constituent aujourd'hui des angles morts du droit (Villani, 2018).
- Par une documentation afin d'évaluer le niveau et la capacité de contrôle existant ou requis dans l'évaluation des systèmes d'IA et de mettre en œuvre une IA fiable (European Commission, 2020).



Source : <https://www.imf.org> /Diebold and Yilmaz (2014)

4

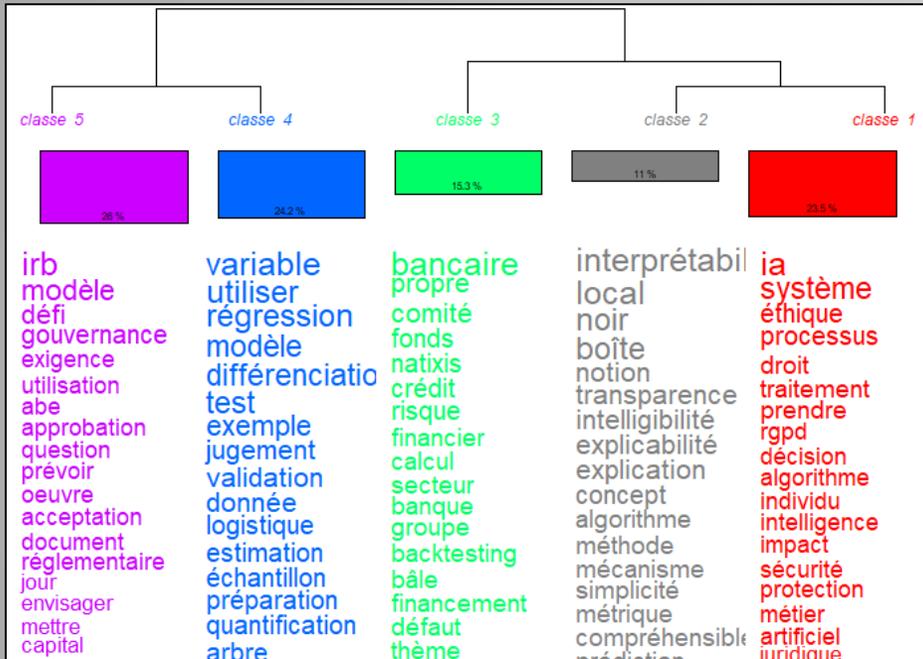
Méthodologie de recherche

Un terrain de recherche existe dans mon entreprise chez NATXIS dans la modélisation interne du risque de crédit en méthode IRBA



Selon nous décidons d'opter pour le raisonnement inductif de (Gauthier, 2005) pour enrichir nos résultats de recherche et de répondre à notre problématique. Les entretiens se sont déroulés sur la base d'une grille d'entretien qui regroupe toutes les questions posées aux participants du groupe en rapport avec notre étude

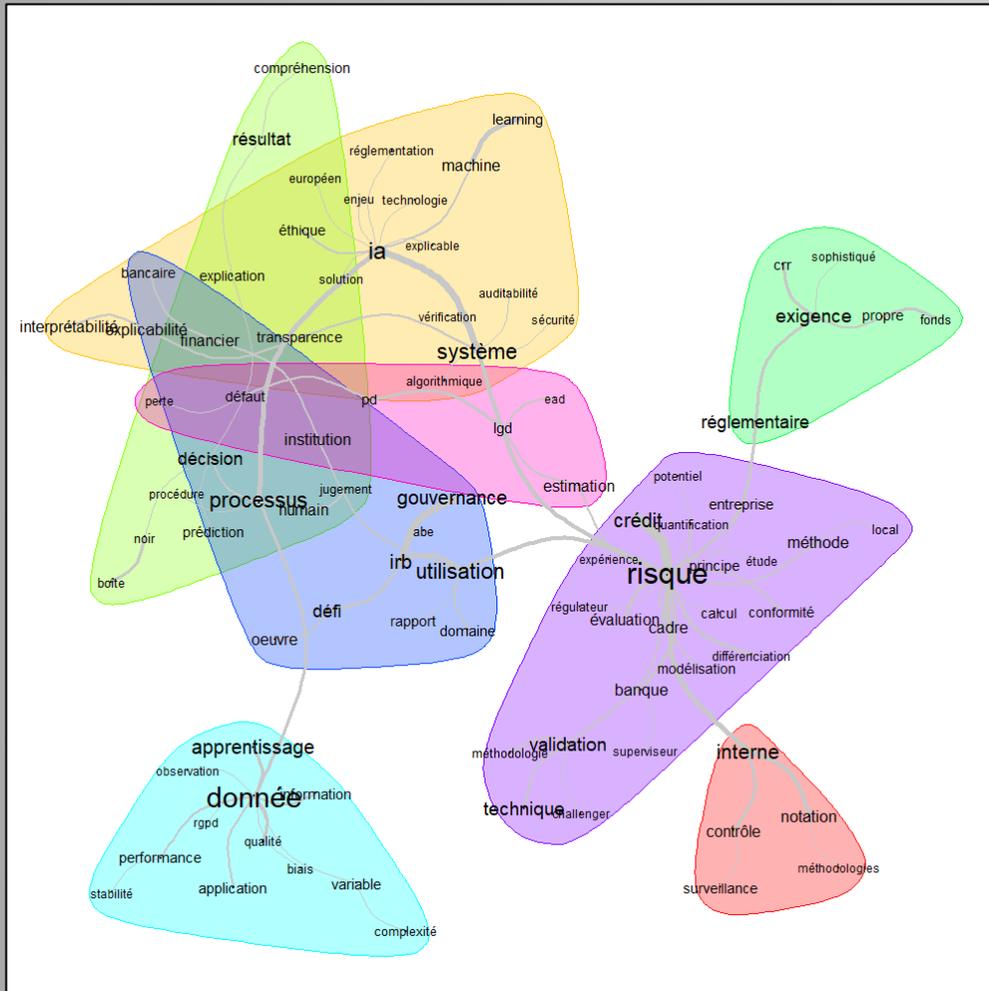
Résultats de la Classification Méthode Reinert - CHD



Dendrogramme des classes, relatif à l'exigence réglementaire et les potentielles solutions du systèmes IA dans le processus du risque de crédit.

Classe 5	Classe 4	Classe 3
<ul style="list-style-type: none"> ▪ Défi d'utiliser le ML dans la modélisation du RC en IRB qui accentue le risque de modèle par la perte financière résultante de l'utilisation de nouveaux modèles. ▪ Renforcement de la gouvernance des risques à appliquer par ce changement, et les exigences de l'ABE sont indispensables ▪ Questionnement sur l'acceptation de la mise en œuvre et les impacts réglementaire 	<ul style="list-style-type: none"> ▪ Variable en référence à la qualité des données en entrée. ▪ Régression logistique et arbre de décision mis en avant dans les choix des modèles du ML. ▪ Aspects pratiques et les expérimentations où les participants des équipes sont impliqués dans le projet pilote du calcul des paramètres du défaut (PD et LGD). 	<ul style="list-style-type: none"> ▪ Comité Bâle III sur l'usage du et de son impact sur les exigences en fonds ▪ Supervision des risques qui a contribué à cette classe sur les aspects du Pilier I ▪ Proposition des exercices de simulations de type « backtesting »
Classe 2	Classe 1	
<ul style="list-style-type: none"> ▪ Interprétabilité et explicabilité des algorithmes, les inquiétudes sur les « boîtes noires ». ▪ Fournir de l'intelligibilité aux métriques. Aussi, ces équipes métiers sont situées en bout de chaîne dans le processus de ▪ Travailler sur des approches locales SHAP ou LIME 	<ul style="list-style-type: none"> ▪ Gestion des algorithmes, les décisions prises contraire à l'éthique et au droit juridique. ▪ Le cadre réglementaire et éthique a un rôle à jouer dans le maintien de la confiance entre les systèmes et les métiers. ▪ les impacts juridiques sur le non-respect de la protection et la sécurité des données dans le cadre du RGPD 	

Analyse des similitudes



Analyse de similitude (6 formes actives les plus présentes par classe)

- L'analyse de similitude permet de représenter la structure du corpus et des relations qu'elle génère, elle met en évidence la connectivité des formes des segments de texte.
- Cela permet des graphes connectés qui modélisent le texte de la manière la plus simple tout en préservant l'architecture principale

une ramification composée de 6 formes qui caractérisent notre corpus :

- Le risque de crédit, les systèmes d'IA, le processus décisionnel, l'exigence réglementaire représentant les thématiques de notre problématique.
- Des thèmes non attendu qui a émergé comme la donnée et la gouvernance des modèles.

6

Solutions proposées

Modèle
« challenger »

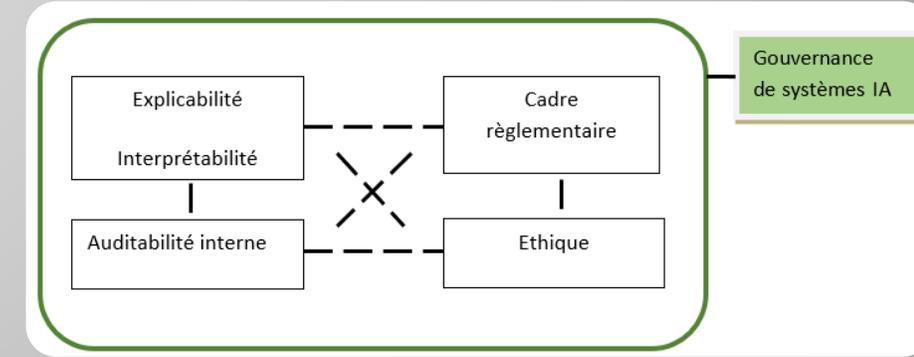
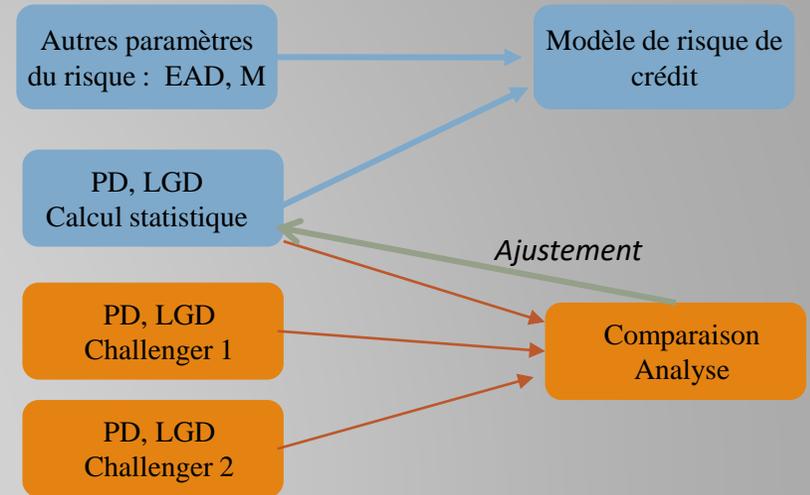
- Comparaison des modèles ML et du modèle statistique courant dans les prédictions des paramètres du défaut PD et LGD.
- Utilisation de la régression logistique et de l'arbre de décision
- Simulation de la probabilité de défaut dans le cadre de demande de financement d'un client institutionnel ou de l'évaluation de risque du crédit durant la vie du contrat dans le cadre de Bâle III.
- Correction ou calibration du modèle statistique de production.

Open Data

- Collecte des données massives et variées en provenance des organismes publics
- Nouveau marché avantageux aux acteurs financiers pour optimiser le potentiel des données en toute sécurité
- Sourcing : INSEE, BCE et FED
- Intégrer les données dans le modèle « challenger » et apporter de nouveaux indices de confiances.

Gouvernance des
algorithme d'IA

- Renforcement de la gouvernance actuelle en suivant les nouvelles préconisations de l'ACPR.
- Formalisation des procédures opérationnelles adaptés à chaque métier.
- Système uniforme de contrôle interne devrait inclure des contrôles de premier niveau par les métiers.
- Surveillance des risques à travers une cartographie des risques.
- Communication et performance entre les différentes dimensions



7 Limites de la recherche

Discussions

- Les machines ne pourront tout faire automatiquement. nous pouvons accélérer significativement les processus avec des compétence une gouvernance holistique, et respectant les recommandations de l'ABE et l'ACPR.
- Si l'humain prend une part trop importante dans les processus, le système ne pourra pas devenir plus intelligent.
- Les tests sur ce périmètre reste expérimentale, nous sommes dans une méthodologie de projection en vue d'implémenter en production un jour, le calcul du modèle IRB avec l'usage des solutions IA.
- Il nous semble un peu prématuré de qualifier l'apprentissage automatique comme une discipline révolutionnaire.

Limites rencontrées

- A ce stade de recherche, le ML n'est pas une discipline révolutionnaire pour la modélisation du risque de crédit, l'émergence des données massives et de la puissance de calcul des machines, devenus des leviers principaux à cette technologie.
- L'utilisation du ML dans la banque n'est pas encore fiabilisée dans notre domaine, et la mise en place des recommandations du superviseur sont lourdes et coûteuses à mettre en place.
- Notre étude se focalise sur échantillon restreint avec 20 interviews auprès des collaborateurs de mon du groupe. Il ne nous ne permet pas de généraliser les conclusions.

- Le travail que nous avons réalisé pourrait être complété et poursuivi sous différents aspects. Il serait pertinent d'étendre cette étude au niveau national ou européen avec le retour d'expériences des autres institutions bancaires.
- Compléter la recherche par les choix empruntés des banques dépendantes d'une réglementation extra-européenne.
- L'utilisation des « clouds » privés issus des fintechs (entreprise de technologie numérique innovante) ouvertes aux banques : un des leviers clés pour transformer les systèmes d'information des entreprises par une augmentation considérable des variables financières pour faire de meilleure prédiction sous le ML.

Merci pour votre écoute