Bienvenue à l'atelier Intelligence Artificielle !

FORUM FINTECH 2021

ACPR
BANQUE DE FRANCE

Travaux passés sur l'IA : la théorie et la pratique

Ateliers expérimentaux (2019-2020)

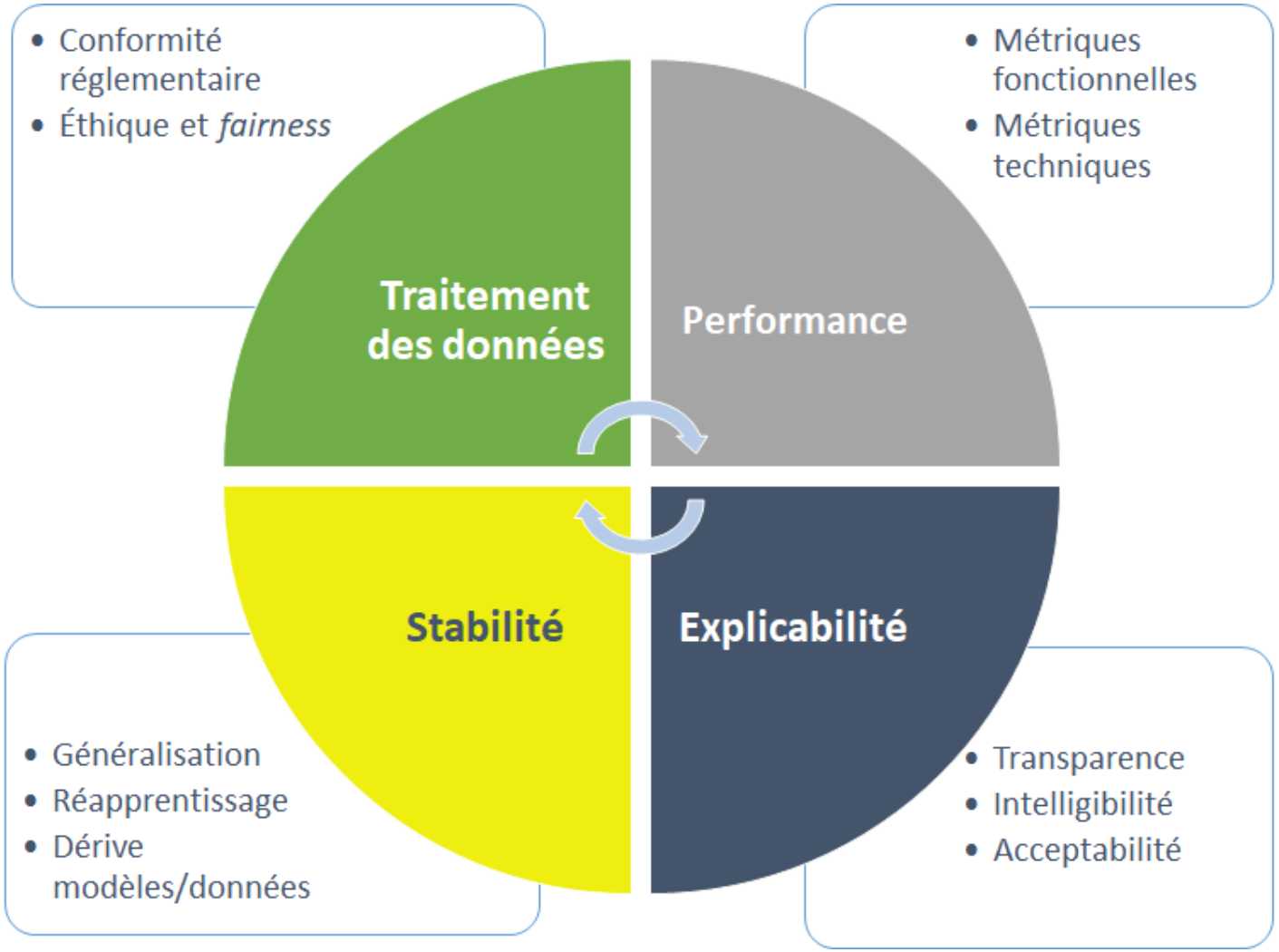
- LCB-FT
- Modèles internes
- Protection du consommateur

Document de discussion et consultation publique (2020)

Webinaires (2020-2021)

Tech Sprint (été 2021)

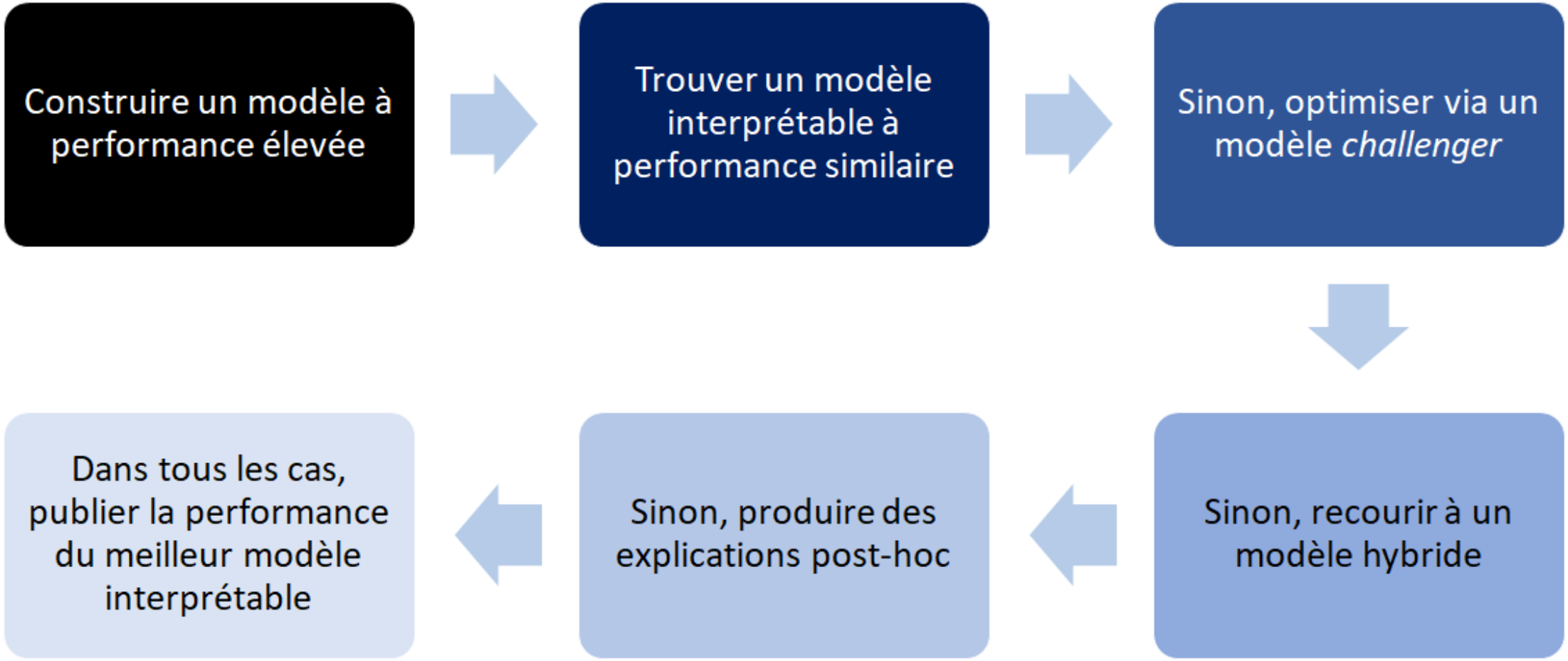
Principes de conception du Machine Learning



Niveaux d'explication

Chemin possible vers des modèles interprétables

"Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead."
Cynthia Rudin (2018)



La boîte noire : un mal parfois nécessaire ?

NLP

Unsupervised Learning

Algorithms that remember: model inversion attacks and data protection law.
Michael Veale et al. (2018)

Vendor Models

NIST Study Shows Face Recognition Experts Perform Better With AI as Partner

Domaine
- automatisation de processus en assurance
- conformité
- fraude
- investissement
- marketing
- modèles de risques
- productivité
- relation client
- sécurité financière
- vente

Usage de l'IA en finance : *réponse à la consultation (décembre 2020)*

Le Tech Sprint « explicabilité » : retour d'expérience et enseignements


tech
SPRINT
08 › 09 juil.21
ACPR | Banque de France


tech
SPRINT
30.06 › 01.07
ACPR | Banque de France
tech
SPRINT
DE L'EXPLICABILITÉ
BANQUE DE FRA
EUROSYSTÈME


tech SPRINT
08 › 09 juil.21
ACPR | Banque de France
Pour rejoindre la session, allez sur : ahaslides.com/TECHSPRINT
Classement final 13 équipes
1 ENSAI-ENSAE ChFr 529p +144
2 Pangea Lam Son Antony Willie Antony Arguirov 414p +124
3 The explAIners Rida 401p +240
4 Just do It(ō) Bob Marley Ben Mouss Rémi 348p +120
5 Akur8 Aymon Robin Guillaume 331p +106
ENSAI-ENSAE
ChFr
529 points

Les modèles de risque de crédit : *pourquoi ce cas d'usage ?*

program. My wife and I filed joint tax returns, live in a community-property state, and have been married for a long time. Yet Apple's black box algorithm thinks I deserve 20x the credit limit she does. No appeals work.

Steve Wozniak ✓
@stevewoz

The same thing happened to us. I got 10x the credit limit. We have no separate bank or credit card accounts or any separate assets. Hard to get to a human for a correction though. It's big tech in 2019.

1:51 AM · 10 nov. 2019

- **Enjeux d'une bonne modélisation**
- **Financement de la consommation**
- **Stabilité financière**
- **Enjeux sociaux** (inclusion financière, non discrimination)

Un défi à 3 étages

Adaptation à une situation d'audit en blackbox

Prise en compte des enjeux métier du risque de crédit

Défi technique d'explicabilité générique du ML

OBJECTIF PRINCIPAL

Faire comprendre le fonctionnement des modèles proposés

L'explication algorithmique consiste à faire comprendre le fonctionnement de l'algorithme et à expliquer pourquoi une prédiction est produite (ou dans le cas de modèles décisionnels, pourquoi une décision est prise).

« Une bonne explication est une explication adaptée à son destinataire.

LE BONUS

Si votre équipe est parvenue à remplir l'objectif principal du Tech Sprint et il qu'il vous reste du temps

Mesurer l'équité algorithmique (ou fairness)

Sur certains modèles, vous pourrez :

définir les biais de nature problématique ou pas (biais de classification ou de prédiction, ou biais statistiques non souhaités déjà présents dans les données) ;

caractériser et quantifier ces biais, par des métriques ou des méthodes explicatives appropriées ;

déterminer dans quelle mesure les biais présents dans les données sont reflétés, voire renforcés, par les modèles de Machine Learning.

NON-OBJECTIFS DU TECH SPRINT

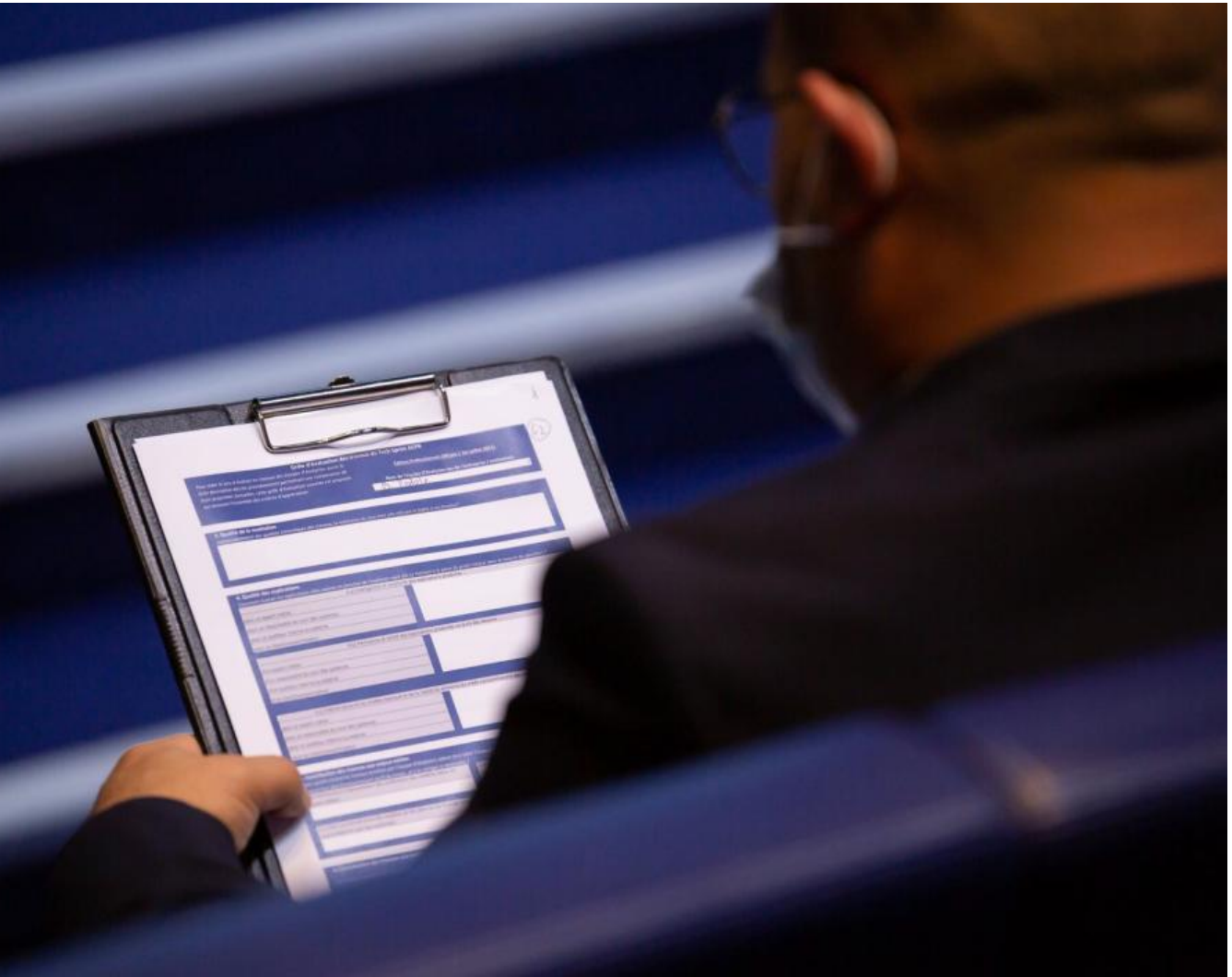
mesurer la performance des modèles

Il ne s'agira pas d'évaluer la performance des modèles prédictifs (par exemple si les estimations de probabilité de défaut à un an sont proches de la réalité une fois l'année écoulée).

mesurer le bien-fondé des prédictions

Il ne s'agira pas non plus de déterminer le bien-fondé des prédictions que les modèles produisent (par exemple si une décision de refus de crédit sur la base d'une prédiction est contestable).

Évaluation des travaux



accomplissements techniques

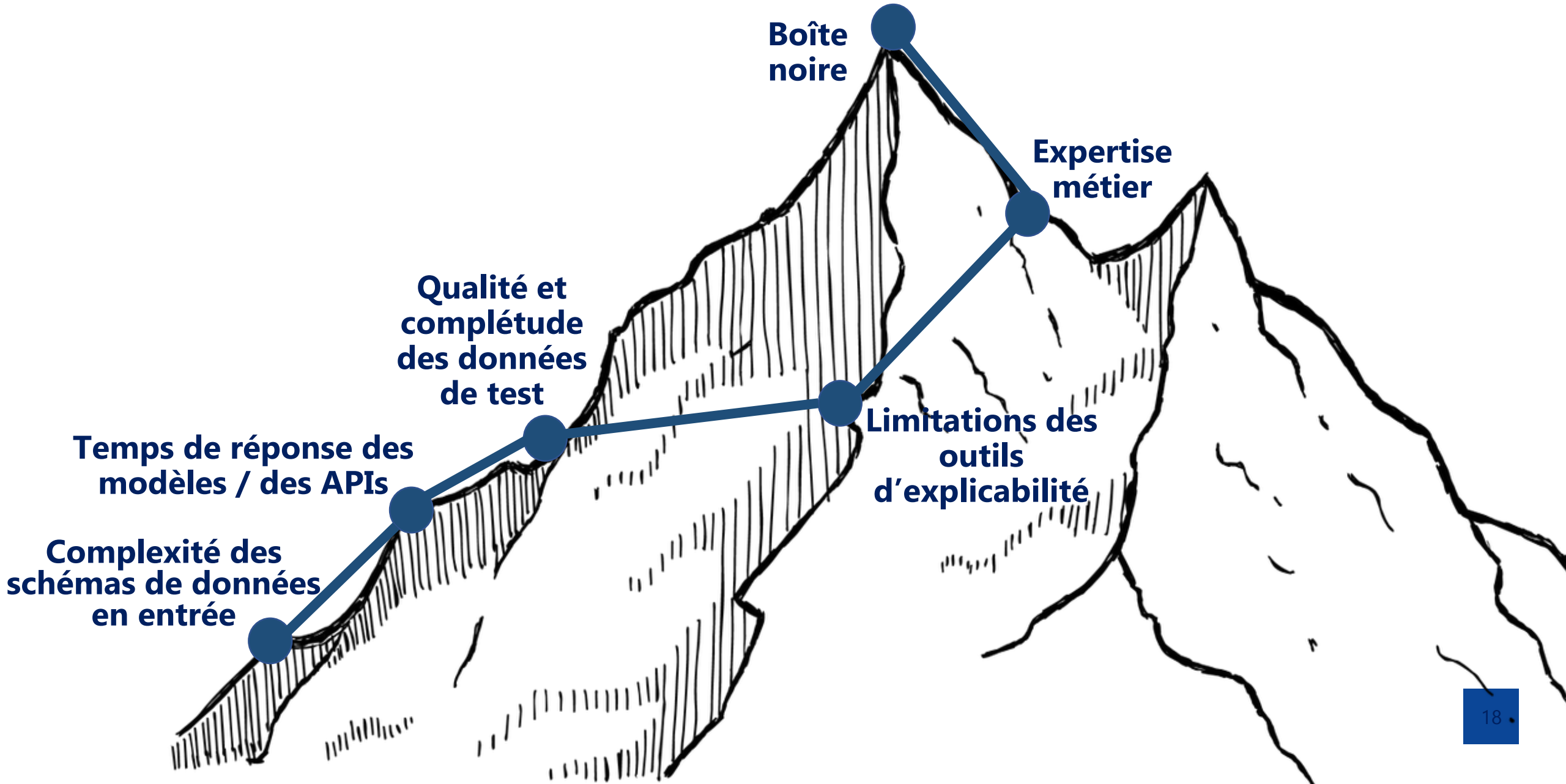
innovation scientifique et méthodologique

caractère clair, **pédagogique** et utile de la restitution

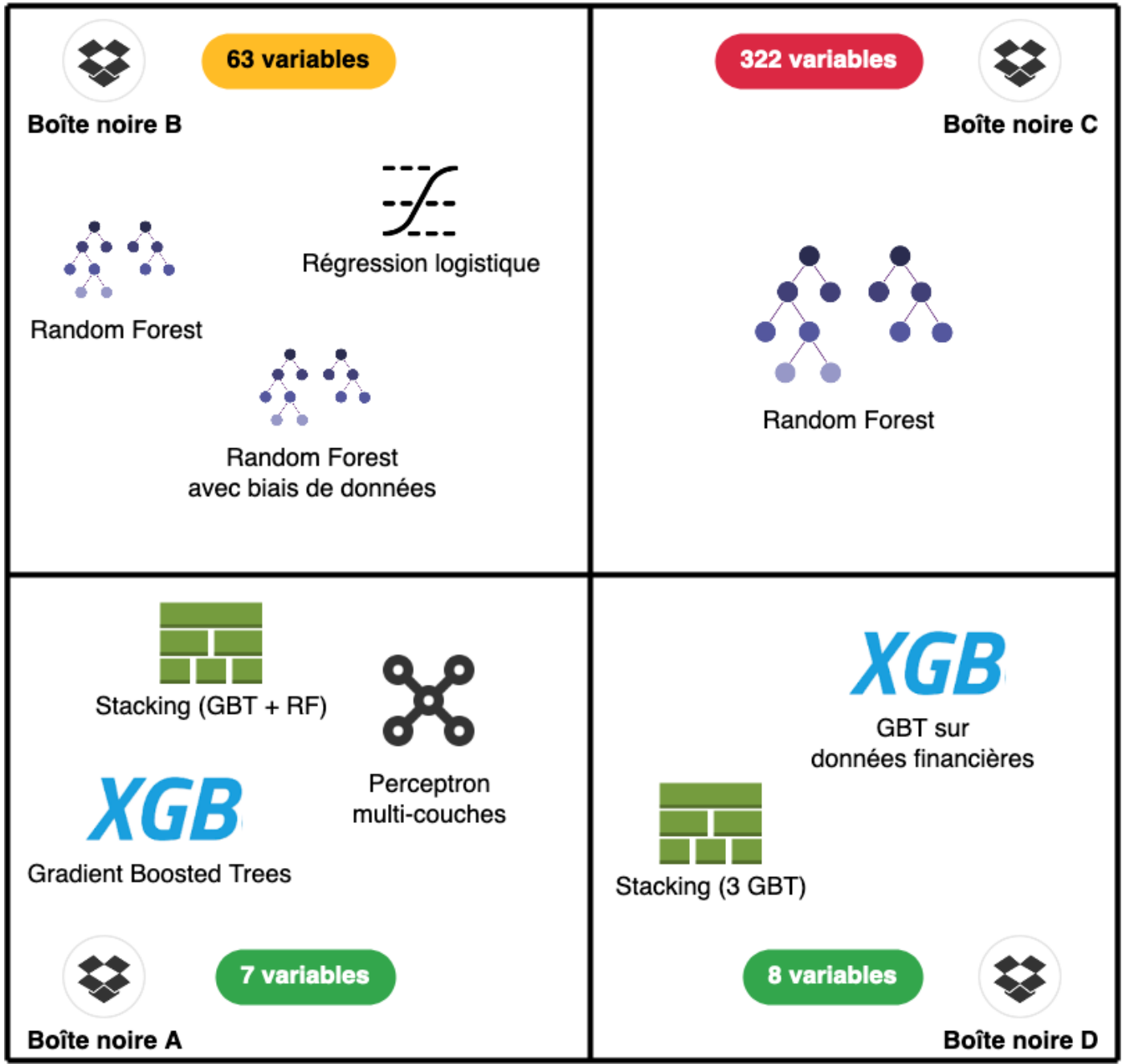
contribution aux **enjeux métier** du risque de crédit

éclairage des **enjeux réglementaires**

Un Sprint semé d'embûches



Dévoilement des boîtes noires



La complexité des schémas de données : *de 322 variables…*

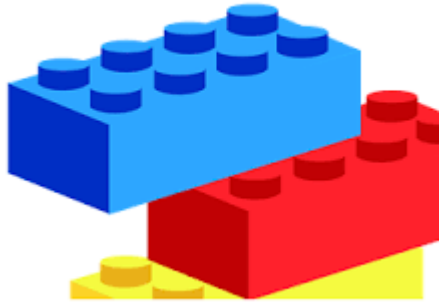
"nb_cav": 1, "mois_ouv": 199801, "SUM_mnt_dec_aut_0": 1600, "AVG_mnt_dep_0": 0, "SUM_mnt_dep_0": 0, "MAX_mnt_dep_0": 0, "AVG_mnt_mvt_cr_0": 3454.85, "MAX_mnt_mvt_cr_0": 3454.85, "AVG_nb_jr_dep_0": 0, "MAX_nb_jr_dep_0": 0, "AVG_nbr_jr_cr_0": 27, "MIN_nbr_jr_cr_0": 27, "AVG_nbr_jr_db_0": 4, "MAX_nbr_jr_db_0": 4, "AVG_nbr_mvt_cr_0": 14, "MIN_nbr_mvt_cr_0": 14, "AVG_nbr_mvt_deb_0": 48, "MAX_nbr_mvt_deb_0": 48, "AVG_sld_fin_mois_0": 1486.58, "SUM_sld_fin_mois_0": 1486.58, "AVG_sld_moy_0": 730, "SUM_sld_moy_0": 730, "AVG_sld_moy_cr_0": 740, "SUM_sld_moy_cr_0": 740, "MIN_sld_moy_cr_0": 740, "AVG_sld_moy_db_0": -10, "SUM_sld_moy_db_0": -10, "MAX_sld_moy_db_0": -10, "SUM_mnt_dec_aut_1": 1600, "AVG_mnt_dep_1": 0, "SUM_mnt_dep_1": 0, "MAX_mnt_dep_1": 0, "AVG_mnt_mvt_cr_1": 7690.51, "MAX_mnt_mvt_cr_1": 7690.51, "AVG_nb_jr_dep_1": 0, "MAX_nb_jr_dep_1": 0, "AVG_nbr_jr_cr_1": 19, "MIN_nbr_jr_cr_1": 19, "AVG_nbr_jr_db_1": 11, "MAX_nbr_jr_db_1": 11, "AVG_nbr_mvt_cr_1": 12, "MIN_nbr_mvt_cr_1": 12, "AVG_nbr_mvt_deb_1": 75, "MAX_nbr_mvt_deb_1": 75, "AVG_sld_fin_mois_1": 2142.17, "SUM_sld_fin_mois_1": 2142.17, "AVG_sld_moy_1": 950, "SUM_sld_moy_1": 950, "AVG_sld_moy_cr_1": 1110, "SUM_sld_moy_cr_1": 1110, "MIN_sld_moy_cr_1": 1110, "AVG_sld_moy_db_1": -160, "SUM_sld_moy_db_1": -160, "MAX_sld_moy_db_1": -160, "SUM_mnt_dec_aut_2": 900, "AVG_mnt_dep_2": 0, "SUM_mnt_dep_2": 0, "MAX_mnt_dep_2": 0, "AVG_mnt_mvt_cr_2": 4526.07, "MAX_mnt_mvt_cr_2": 4526.07, "AVG_nb_jr_dep_2": 0, "MAX_nb_jr_dep_2": 0, "AVG_nbr_jr_cr_2": 4, "MIN_nbr_jr_cr_2": 4, "AVG_nbr_jr_db_2": 27, "MAX_nbr_jr_db_2": 27, "AVG_nbr_mvt_cr_2": 21, "MIN_nbr_mvt_cr_2": 21, "AVG_nbr_mvt_deb_2": 56, "MAX_nbr_mvt_deb_2": 56, "AVG_sld_fin_mois_2": 1340.38, "SUM_sld_fin_mois_2": 1340.38, "AVG_sld_moy_2": -260, "SUM_sld_moy_2": -260, "AVG_sld_moy_cr_2": 90, "SUM_sld_moy_cr_2": 90, "MIN_sld_moy_cr_2": 90, "AVG_sld_moy_db_2": -360, "SUM_sld_moy_db_2": -360, "MAX_sld_moy_db_2": -360, "SUM_mnt_dec_aut_3": 900, "AVG_mnt_dep_3": 0, "SUM_mnt_dep_3": 0, "MAX_mnt_dep_3": 0, "AVG_mnt_mvt_cr_3": 2698.13, "MAX_mnt_mvt_cr_3": 2698.13, "AVG_nb_jr_dep_3": 0, "MAX_nb_jr_dep_3": 0, "AVG_nbr_jr_cr_3": 11, "MIN_nbr_jr_cr_3": 11, "AVG_nbr_jr_db_3": 19, "MAX_nbr_jr_db_3": 19, "AVG_nbr_mvt_cr_3": 6, "MIN_nbr_mvt_cr_3": 6, "AVG_nbr_mvt_deb_3": 53, "MAX_nbr_mvt_deb_3": 53, "AVG_sld_fin_mois_3": 538.82, "SUM_sld_fin_mois_3": 538.82, "AVG_sld_moy_3": -190, "SUM_sld_moy_3": -190, "AVG_sld_moy_cr_3": 150, "SUM_sld_moy_cr_3": 150, "MIN_sld_moy_cr_3": 150, "AVG_sld_moy_db_3": -350, "SUM_sld_moy_db_3": -350, "MAX_sld_moy_db_3": -350, "SUM_mnt_dec_aut_4": 900, "AVG_mnt_dep_4": 0, "SUM_mnt_dep_4": 0, "MAX_mnt_dep_4": 0, "AVG_mnt_mvt_cr_4": 1763.72, "MAX_mnt_mvt_cr_4": 1763.72, "AVG_nb_jr_dep_4": 0, "MAX_nb_jr_dep_4": 0, "AVG_nbr_jr_cr_4": 15, "MIN_nbr_jr_cr_4": 15, "AVG_nbr_jr_db_4": 16, "MAX_nbr_jr_db_4": 16, "AVG_nbr_mvt_cr_4": 7, "MIN_nbr_mvt_cr_4": 7, "AVG_nbr_mvt_deb_4": 59, "MAX_nbr_mvt_deb_4": 59, "AVG_sld_fin_mois_4": 699.26, "SUM_sld_fin_mois_4": 699.26, "AVG_sld_moy_4": 90, "SUM_sld_moy_4": 90, "AVG_sld_moy_cr_4": 250, "SUM_sld_moy_cr_4": 250, "MIN_sld_moy_cr_4": 250, "AVG_sld_moy_db_4": -160, "SUM_sld_moy_db_4": -160, "MAX_sld_moy_db_4": -160, "SUM_mnt_dec_aut_5": 900, "AVG_mnt_dep_5": 0, "SUM_mnt_dep_5": 0, "MAX_mnt_dep_5": 0, "AVG_mnt_mvt_cr_5": 2047.25, "MAX_mnt_mvt_cr_5": 2047.25, "AVG_nb_jr_dep_5": 0, "MAX_nb_jr_dep_5": 0, "AVG_nbr_jr_cr_5": 31, "MIN_nbr_jr_cr_5": 31, "AVG_nbr_jr_db_5": 0, "MAX_nbr_jr_db_5": 0, "AVG_nbr_mvt_cr_5": 5, "MIN_nbr_mvt_cr_5": 5, "AVG_nbr_mvt_deb_5": 52, "MAX_nbr_mvt_deb_5": 52, "AVG_sld_fin_mois_5": 834.04, "SUM_sld_fin_mois_5": 834.04, "AVG_sld_moy_5": 590, "SUM_sld_moy_5": 590, "AVG_sld_moy_cr_5": 590, "SUM_sld_moy_cr_5": 590, "MIN_sld_moy_cr_5": 590, "AVG_sld_moy_db_5": 0, "SUM_sld_moy_db_5": 0, "MAX_sld_moy_db_5": 0, "SUM_mnt_dec_aut_6": 900, "AVG_mnt_dep_6": 0, "SUM_mnt_dep_6": 0, "MAX_mnt_dep_6": 0, "AVG_mnt_mvt_cr_6": 3907.49, "MAX_mnt_mvt_cr_6": 3907.49, "AVG_nb_jr_dep_6": 0, "MAX_nb_jr_dep_6": 0, "AVG_nbr_jr_cr_6": 13, "MIN_nbr_jr_cr_6": 13, "AVG_nbr_jr_db_6": 17, "MAX_nbr_jr_db_6": 17, "AVG_nbr_mvt_cr_6": 11, "MIN_nbr_mvt_cr_6": 11, "AVG_nbr_mvt_deb_6": 60, "MAX_nbr_mvt_deb_6": 60, "AVG_sld_fin_mois_6": 1506.39, "SUM_sld_fin_mois_6": 1506.39, "AVG_sld_moy_6": 10, "SUM_sld_moy_6": 10, "AVG_sld_moy_cr_6": 290, "SUM_sld_moy_cr_6": 290, "MIN_sld_moy_cr_6": 290, "AVG_sld_moy_db_6": -270, "SUM_sld_moy_db_6": -270, "MAX_sld_moy_db_6": -270, "nb_credit": 1, "nb_ppi": null, "nb_conso": 1, "AVG_mnt_prt_0": null, "SUM_mnt_prt_0": null, "MAX_mnt_prt_0": null, "AVG_mnt_prt_1": null, "SUM_mnt_prt_1": null, "MAX_mnt_prt_1": null, "AVG_mnt_prt_2": null, "SUM_mnt_prt_2": null, "MAX_mnt_prt_2": null, "AVG_mnt_prt_3": null, "SUM_mnt_prt_3": null, "MAX_mnt_prt_3": null, "AVG_mnt_prt_4": null, "SUM_mnt_prt_4": null, "MAX_mnt_prt_4": null, "AVG_mnt_prt_5": null, "SUM_mnt_prt_5": null, "MAX_mnt_prt_5": null, "AVG_mnt_prt_6": null, "SUM_mnt_prt_6": null, "MAX_mnt_prt_6": null, "AVG_mnt_imp_0": 0, "SUM_mnt_imp_0": 0, "MAX_mnt_imp_0": 0, "AVG_mnt_imp_1": 0, "SUM_mnt_imp_1": 0, "MAX_mnt_imp_1": 0, "AVG_mnt_imp_2": null, "SUM_mnt_imp_2": null, "MAX_mnt_imp_2": null, "AVG_mnt_imp_3": null, "SUM_mnt_imp_3": null, "MAX_mnt_imp_3": null, "AVG_mnt_imp_4": null, "SUM_mnt_imp_4": null, "MAX_mnt_imp_4": null, "AVG_mnt_imp_5": null, "SUM_mnt_imp_5": null, "MAX_mnt_imp_5": null, "AVG_mnt_imp_6": null, "SUM_mnt_imp_6": null, "MAX_mnt_imp_6": null, "AVG_enc_fin_mois_0": 4870, "SUM_enc_fin_mois_0": 4870, "min_enc_fin_mois_0": 4870, "AVG_enc_fin_mois_1": 5000, "SUM_enc_fin_mois_1": 5000, "min_enc_fin_mois_1": 5000, "AVG_enc_fin_mois_2": null, "SUM_enc_fin_mois_2": null, "min_enc_fin_mois_2": null, "AVG_enc_fin_mois_3": null, "SUM_enc_fin_mois_3": null, "min_enc_fin_mois_3": null, "AVG_enc_fin_mois_4": null, "SUM_enc_fin_mois_4": null, "min_enc_fin_mois_4": null, "AVG_enc_fin_mois_5": null, "SUM_enc_fin_mois_5": null, "min_enc_fin_mois_5": null, "AVG_enc_fin_mois_6": null, "SUM_enc_fin_mois_6": null, "min_enc_fin_mois_6": null, "SUM_nb_ech_restant": null, "MAX_nb_ech_restant": 999999, "sum_nb_imp": 0, "MAX_nb_imp": 0, "AVG_epargne_fin_mois_0": 1120, "SUM_epargne_fin_mois_0": 1120, "MAX_epargne_fin_mois_0": 1120, "AVG_epargne_fin_mois_1": 0, "SUM_epargne_fin_mois_1": 0, "MAX_epargne_fin_mois_1": 0, "AVG_epargne_fin_mois_2": 60, "SUM_epargne_fin_mois_2": 60, "MAX_epargne_fin_mois_2": 60, "AVG_epargne_fin_mois_3": 50, "SUM_epargne_fin_mois_3": 50, "MAX_epargne_fin_mois_3": 50, "AVG_epargne_fin_mois_4": 30, "SUM_epargne_fin_mois_4": 30, "MAX_epargne_fin_mois_4": 30, "AVG_epargne_fin_mois_5": 20, "SUM_epargne_fin_mois_5": 20, "MAX_epargne_fin_mois_5": 20, "AVG_epargne_fin_mois_6": 0, "SUM_epargne_fin_mois_6": 0, "MAX_epargne_fin_mois_6": 0, "sexe": "F", "annee_naissance": 1982, "departement_res": null, "anciennete_rel": "2-5 ans", "SUM_mnt_imp_sem_1": null, "SUM_mnt_imp_sem_2": 999999, "AVG_nb_jr_dep_sem_1": 0, "AVG_nb_jr_dep_sem_2": 0, "MIN_nbr_jr_cr_sem_1": 11.3333333333333, "MIN_nbr_jr_cr_sem_2": 19.6666666666667, "AVG_sld_moy_sem_1": 166.666666666667, "AVG_sld_moy_sem_2": 230, "AVG_sld_moy_cr_sem_1": 450, "AVG_sld_moy_cr_sem_2": 376.666666666667, "AVG_sld_fin_mois_sem_1": 1340.45666666667, "AVG_sld_fin_mois_sem_2": 1013.23, "SUM_mnt_dec_aut_sem_1": 1133.33333333333, "SUM_mnt_dec_aut_sem_2": 900, "MAX_nbr_mvt_deb_sem_1": 61.3333333333333, "MAX_nbr_mvt_deb_sem_2": 57, "AVG_mnt_mvt_cr_sem_1": 4971.57, "AVG_mnt_mvt_cr_sem_2": 2572.82, "AVG_nbr_mvt_cr_sem_1": 13, "AVG_nbr_mvt_cr_sem_2": 7.66666666666667, "AVG_sld_moy_db_sem_1": -290, "AVG_sld_moy_db_sem_2": -143.333333333333, "SUM_mnt_dep_sem_1": 0, "SUM_mnt_dep_sem_2": 0, "AVG_mnt_prt_sem_1": null, "AVG_mnt_prt_sem_2": null, "AVG_epargne_fin_mois_sem_1": 36.6666666666667, "AVG_epargne_fin_mois_sem_2": 16.6666666666667, "age": 39, "SUM_mnt_imp_0_moy": 0, "AVG_nb_jr_dep_0_moy": 0, "AVG_mnt_mvt_cr_0_moy": 3454.85, "AVG_sld_moy_0_moy": 730, "SUM_mnt_dep_0_moy": 0, "MAX_nbr_mvt_deb_0_moy": 48, "SUM_mnt_imp_sem_2_moy": 0.856129057530097, "AVG_sld_moy_db_0_moy": -10, "AVG_nbr_mvt_cr_sem_1_moy": 13, "nb_conso_moy": 1, "SUM_mnt_dep_sem_1_moy": 0, "MIN_nbr_jr_cr_sem_1_moy": 11.3333333333333, "anciennete_rel_moy": "2-5 ans", "age_moy": 39

...à 14 variables utilisées par le modèle !


```
"AVG_nb_jr_dep_0": 0,
"AVG_sld_moy_0_moy": 730,
"SUM_mnt_imp_0_moy": 0,
"age": 39,
"anciennete_rel_moy": "2-5 ans",
"AVG_mnt_mvt_cr_0_moy": 3454.85,
"MIN_nbr_jr_cr_sem_1_moy": 11.3333333333333,
"MAX_nbr_mvt_deb_0_moy": 48,
"AVG_sld_moy_db_0_moy": -10,
"AVG_nbr_mvt_cr_sem_1_moy": 13,
"SUM_mnt_dep_sem_1_moy": 0,
"SUM_mnt_imp_sem_2_moy": 0.856129057530097,
"SUM_mnt_dep_0_moy": 0,
"nb_conso_moy": 1
```

Principaux enseignements

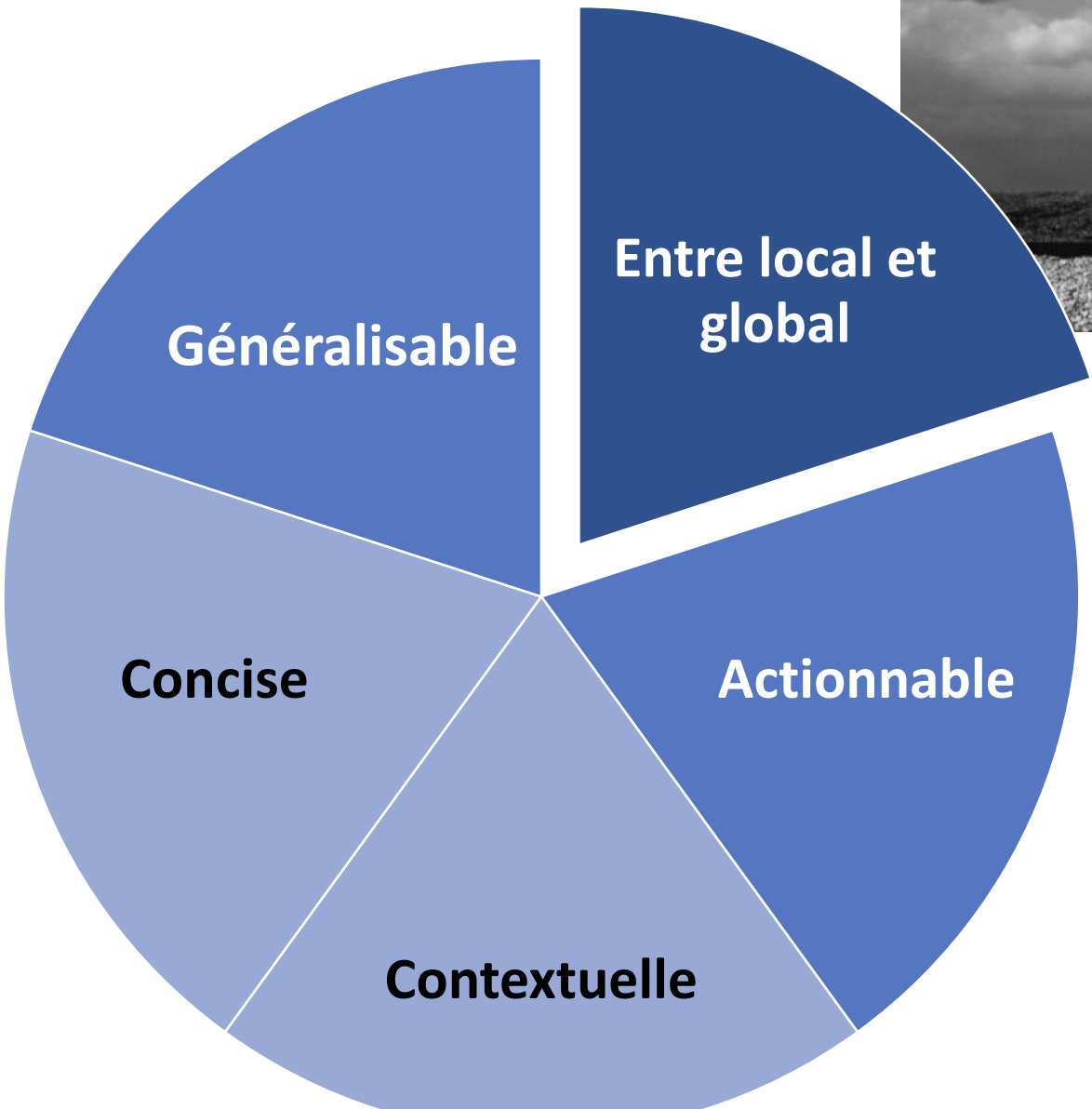
- Importance de la **pluridisciplinarité** et de la polyvalence
- La question du "***build vs. buy***" se pose aussi en ML



- **R&D avancée** dans le secteur
  - Shapash (MAIF)
  - Active Coalition Values (QuantMetry)
  - Skope-rules (BPCE)
- **Simplicité ≠ explicabilité**

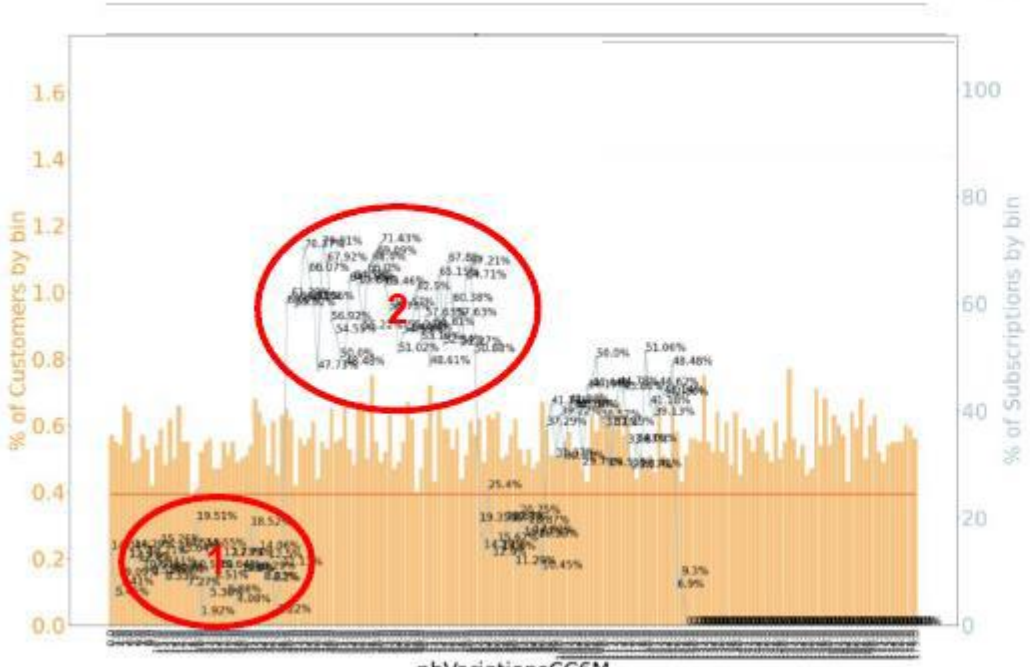
> “L’approche pure boîte noire a montré ses limites et la construction de modèle de ML doit être itérative, agile, multidisciplinaire et documentée ; les explications doivent se faire en présence d’experts métiers et de data science.
>
> **-Un participant au Tech Sprint**

Points d'attention ressortis du Tech Sprint pour une bonne explication de modèle



Entre le local et le global : *les Personae*

Persona 1	Persona 2	Persona 3
Julien	Nicolas	Isis
- 20 ans - locataire - célibataire	- 35 ans - propriétaire - marié	- 50 ans - locataire - mariée



Nos conclusions pour Julien :

Nous mesurons l'impact de l'activité sur le compte courant sur la prédiction.

Si peu d'activité (1) : peu de Julien font défaut.

Au delà (2) : les Julien font plus souvent défaut

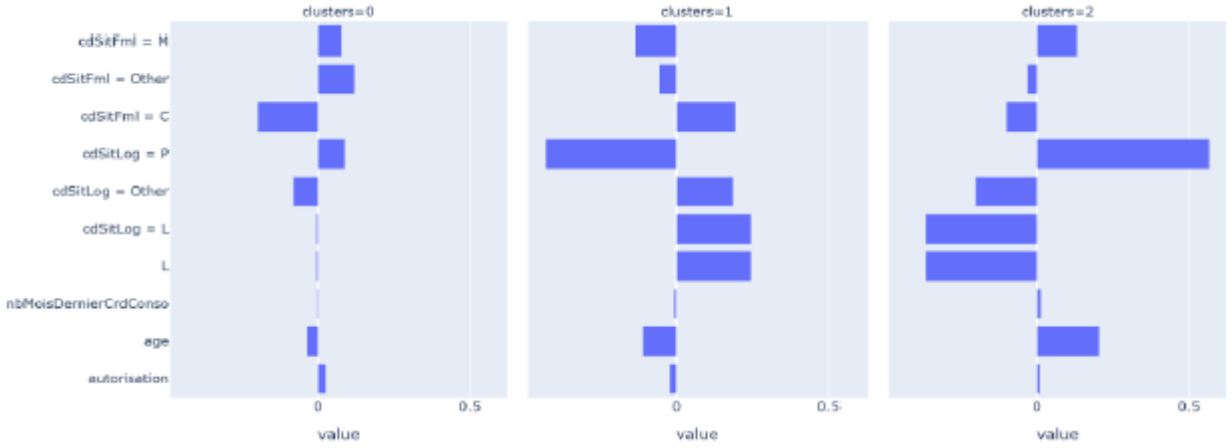
Entre le local et le global : *clustering de valeurs SHAP*

Explications régionales

Echelle globale
les effets importants en
“général”

Echelle régionale
segment de clientèle

Echelle locale
centrée sur un client



Profils-types

Groupe à score faible : individu non célibataire

Groupe à score élevé : individu jeune et non propriétaire

Groupe à score modéré : individu plus âgé et propriétaire

Concision des explications

- **Regroupement de variables** par comportement ou par sémantique

Catégorie 1 :

Données financières :

- Revenus mensuels
- Loyers et mensualités de crédits
- Nombre de crédits en cours
- Situation de logement

Catégorie 2 :

Données sociodémographiques :

- Âge
- Situation maritale
- Situation de logement
- Situation professionnelle

Catégorie 3 :

Données contact/pro :

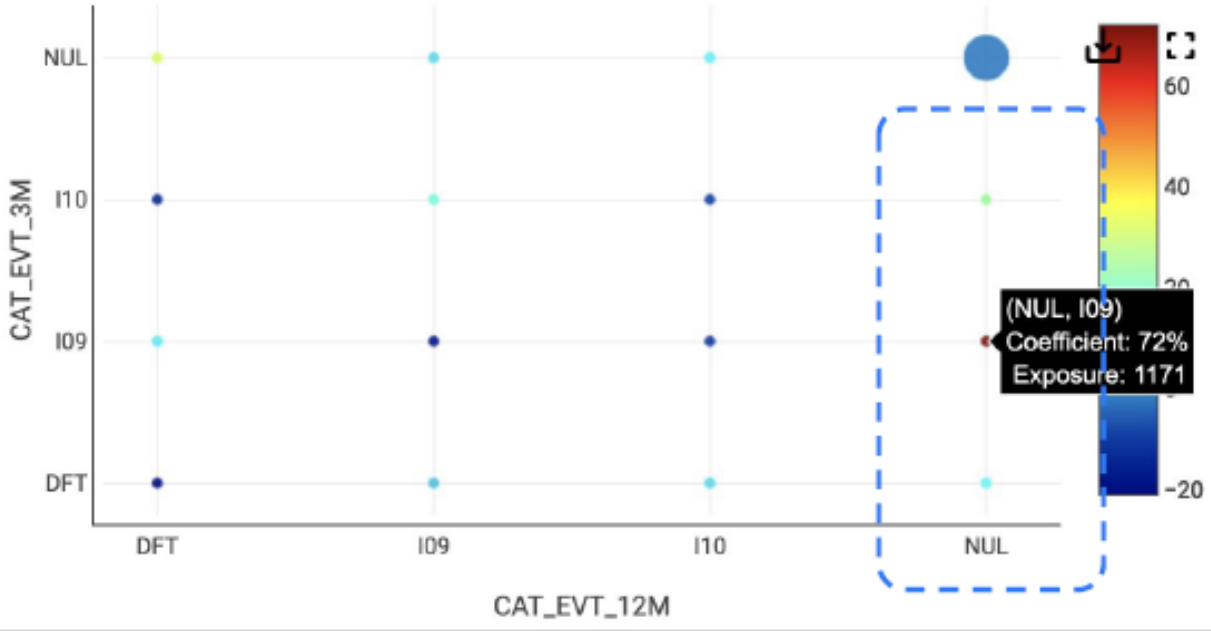
- Domaine d'adresse mail
- Revenus mensuels
- Situation professionnelle

- **Définition** : « à adapter à chaque fois en fonction des parties prenantes »
- Arbitrage probabiliste **concision / robustesse**
- Convergence vers la notion de « ***chunk*** » **cognitif**

Pouvoir de généralisation des explications

Une explication doit-elle rester valide sur :

- le périmètre formel de validité de l'explication ou au-delà ?
- des points aberrants dans les données d'apprentissage (durée de contrat <0 si échéancier modifié) ?
- des données avec certains attributs manquants?
- des données anormales mais plausibles ?
- des données peu plausibles (étudiant de 65 ans) ?



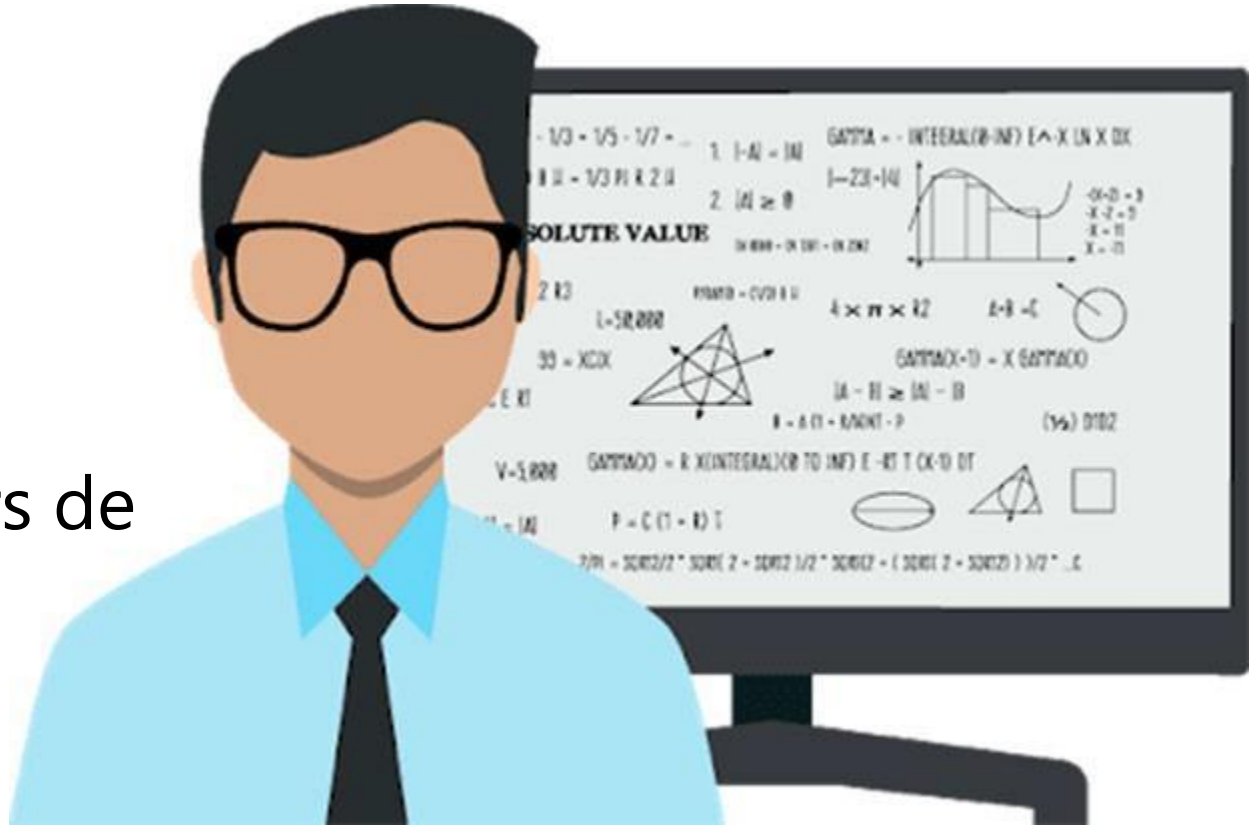
Événements impossibles !

Exemples d'explications produites pour le data scientist

Objectifs : comprendre la nature du modèle, ses performances, le traitement des valeurs manquantes

Modèle de substitution :
arbre relationnel complet

Méthode post-hoc : SDP local avec probabilité des estimations + valeurs de Shapley avec incertitudes



Exemples d'explications produites pour l'expert métier

Objectifs : comprendre ce qui influence la décision, aider le client à éviter de faire défaut

Modèle de substitution : les règles les plus sûres

Méthode post-hoc : règles métier générales (*Skope rules*) + valeurs de SHAP locales

Exemples d'explications produites pour l'auditeur

Objectifs : catégoriser les clients, comprendre l'ensemble du processus et l'objectif du modèle, ses limites, comprendre si le modèle fait ce pour quoi il est conçu

Modèle de substitution : difficile à définir, encore plus à entraîner (surtout en audit externe)

Méthode post-hoc : explications globales + focus sur les régions de « stress »

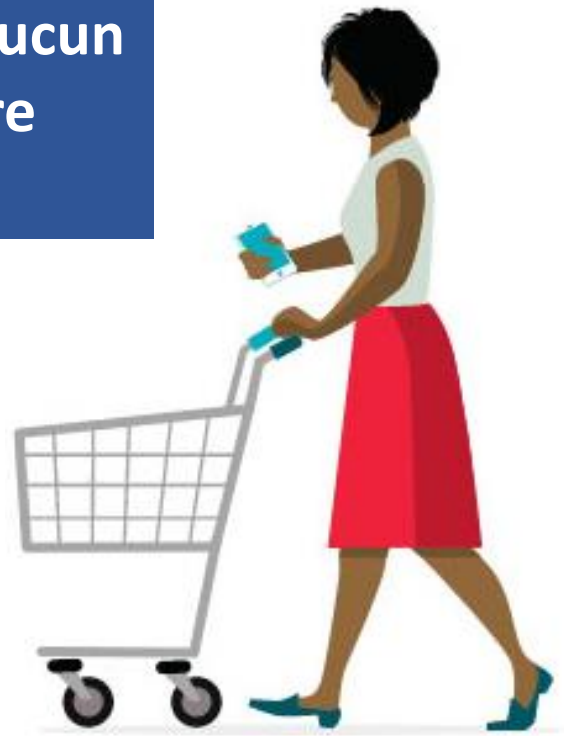
Exemples d'explications produites pour le consommateur

Objectifs : comprendre pourquoi la décision a été prise, comment son comportement influe sur l'algorithme, etc.

Modèle de substitution : explication en langage naturel sur la base des règles les plus sûres

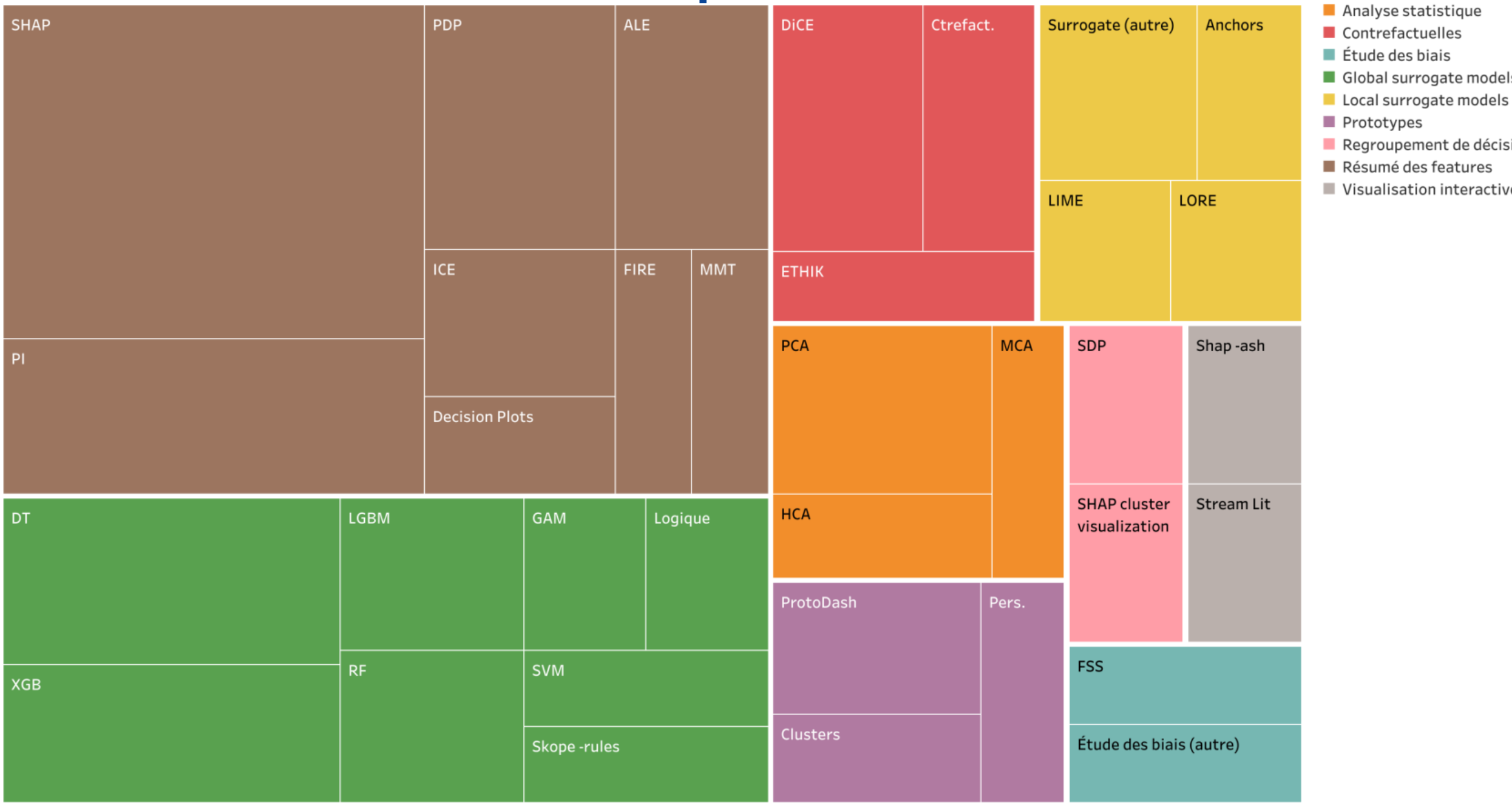
« L'avis est favorable car vous souhaitez racheter ou reprendre un prêt immobilier résidentiel et aucun écriture n'a été écartée sur votre CAV pour les 3 derniers mois. »

Méthode post-hoc : exemples contrefactuels + importance locale des variables

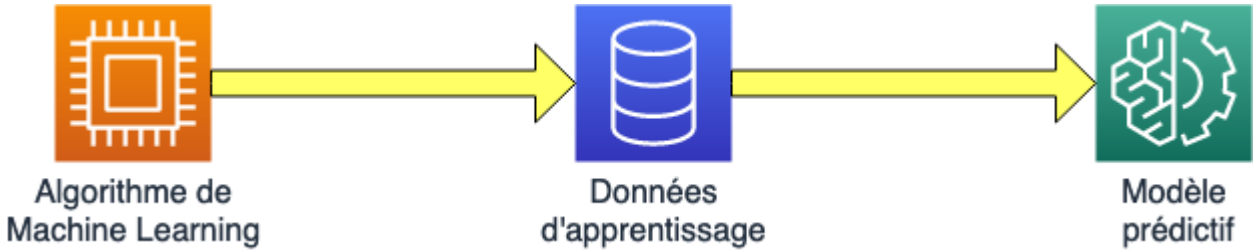


Méthodes explicatives utilisées

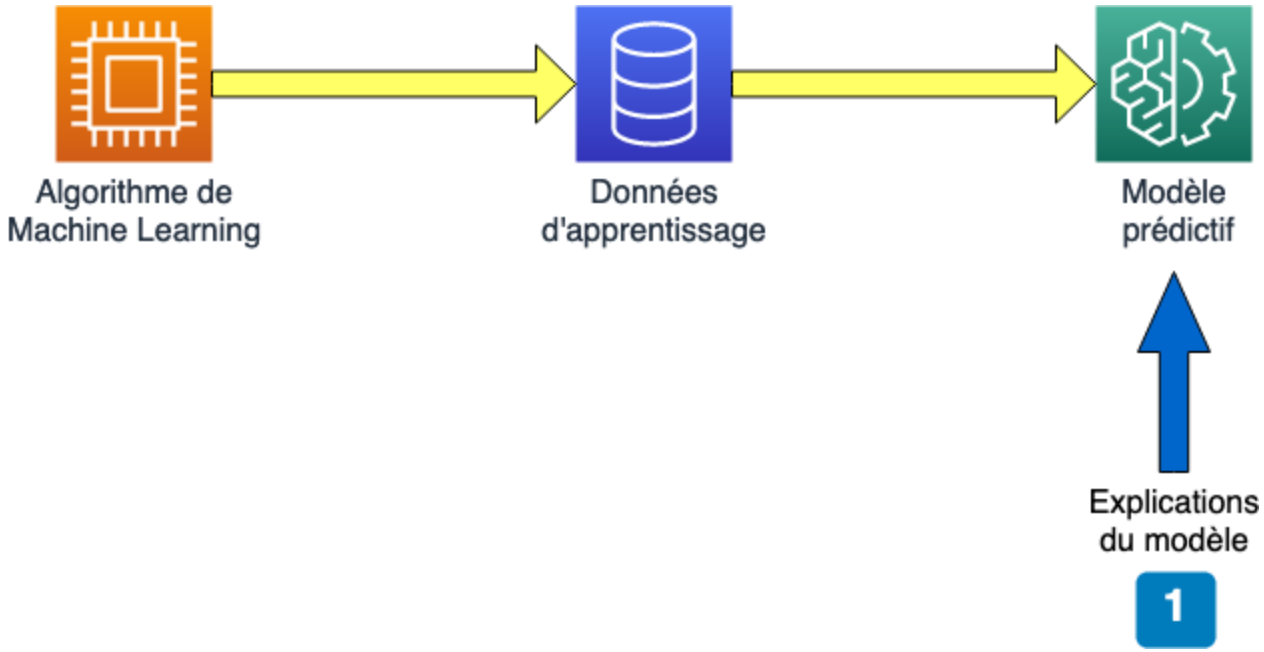
Méthodes explicatives utilisées



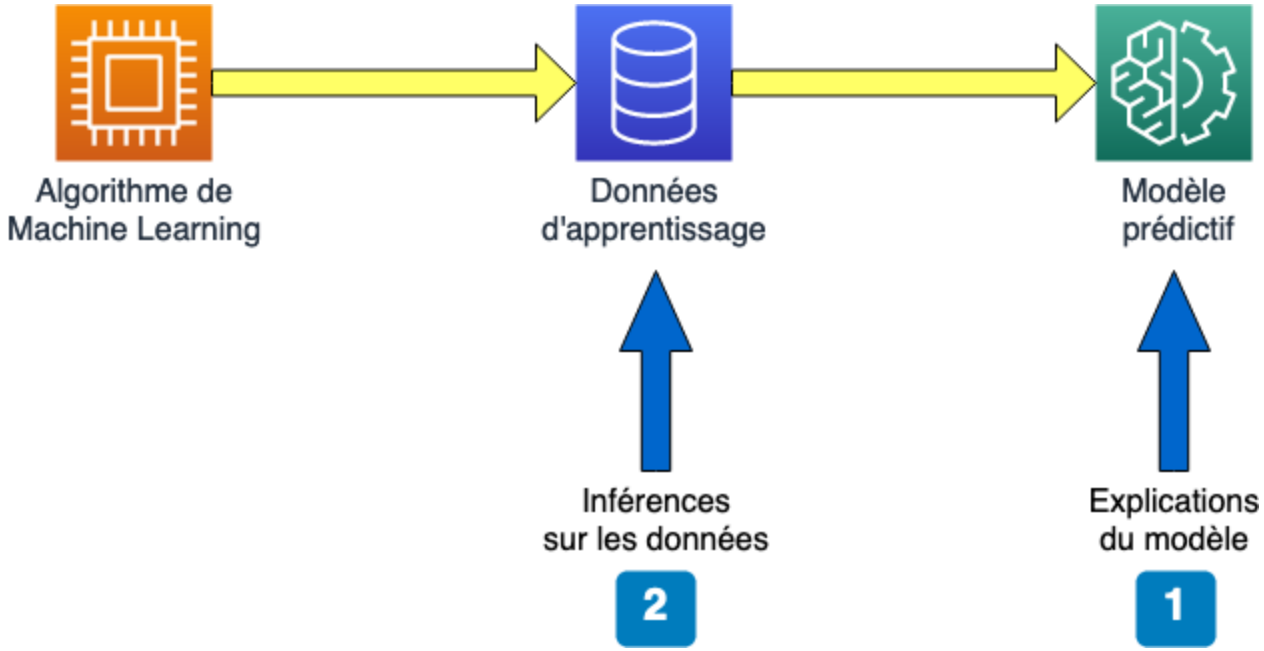
Quelle était la question ?



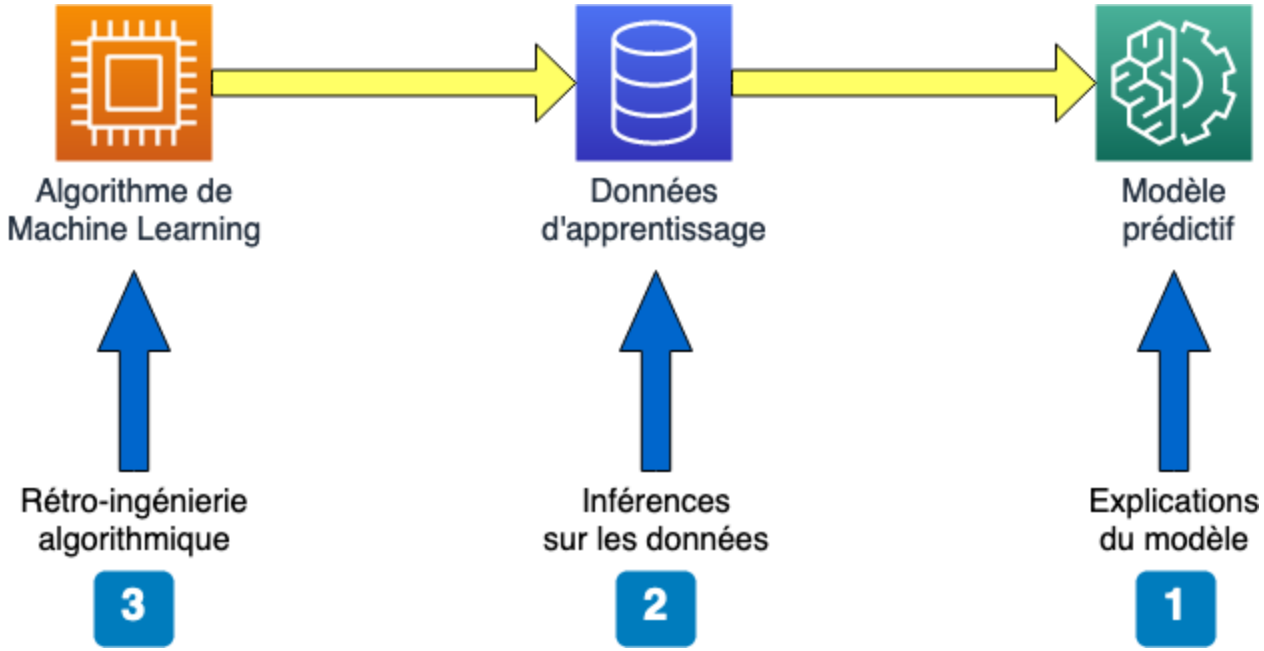
Quelle était la question ?



Quelle était la question ?



Quelle était la question ?



Quelle était la question ?

Donner à comprendre autant que possible :

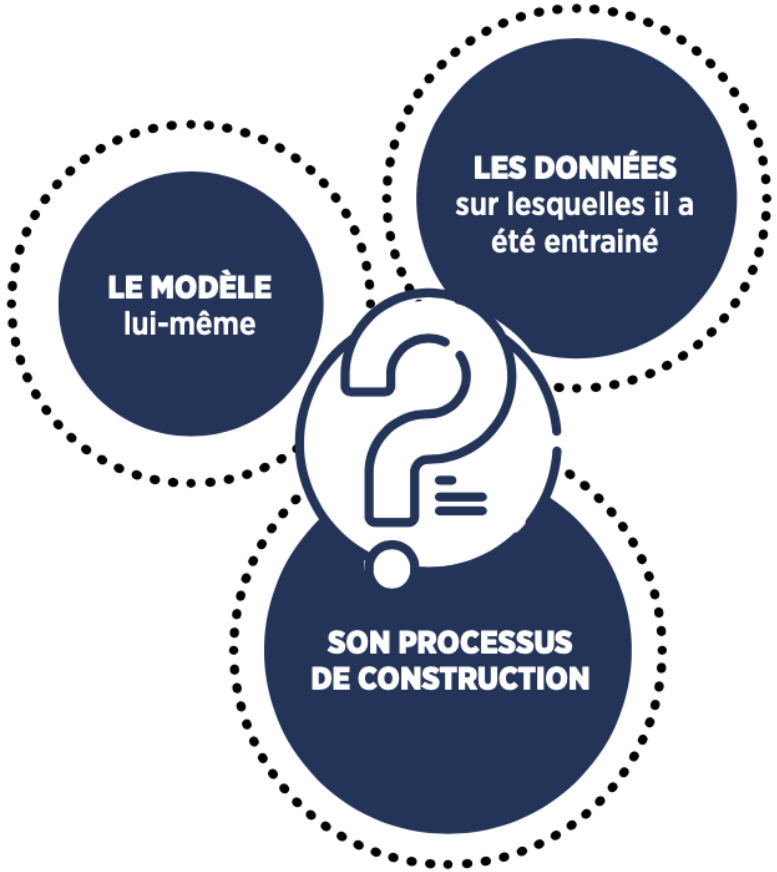
1. **Le modèle lui-même**

2. **Les données sur lesquelles il a été entraîné**

   A savoir : volumétrie, caractéristiques statistiques, anomalies, points ou sous-populations d'intérêt, etc.

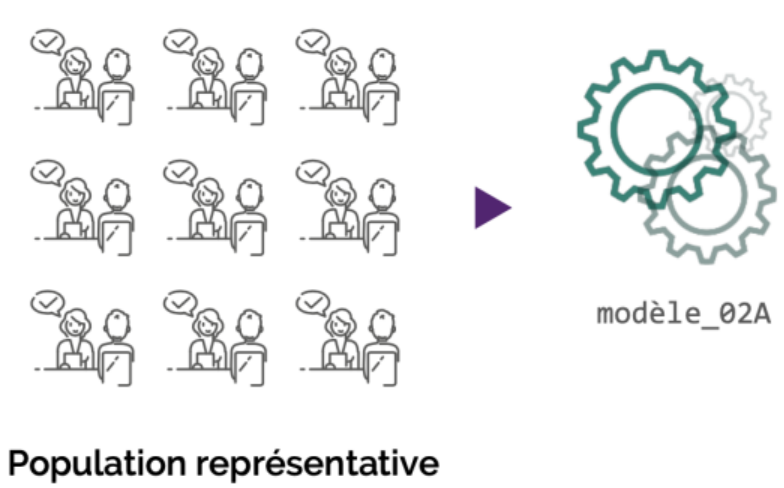
3. **Son processus de construction**

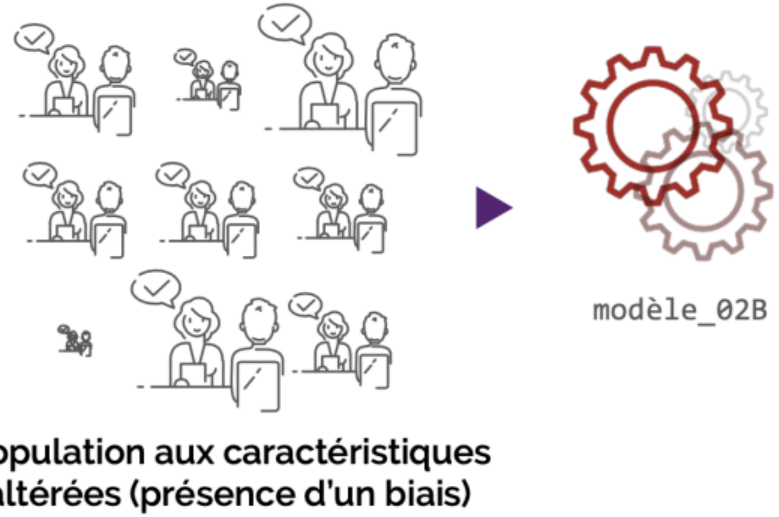
   C'est l'esprit du reverse engineering : inférer non seulement la classe d'algorithme de ML, mais aussi ses hyperparamètres et autres éléments de configuration, toute particularité du modèle donné, mais idéalement aussi le langage de programmation dans lequel il a été implémenté...



Au-delà du modèle :
(2) expliquer les données d'entraînement

- Surtout du calcul d'importance des variables prédictives
- Peu d'inférence statistique (distribution, outliers) et uniquement heuristique
- Exercice de comparaison de modèles difficile (surpondération des clients en défaut sans incident bancaire préalable)





Étape ultime :
(3) expliquer le processus de construction du modèle

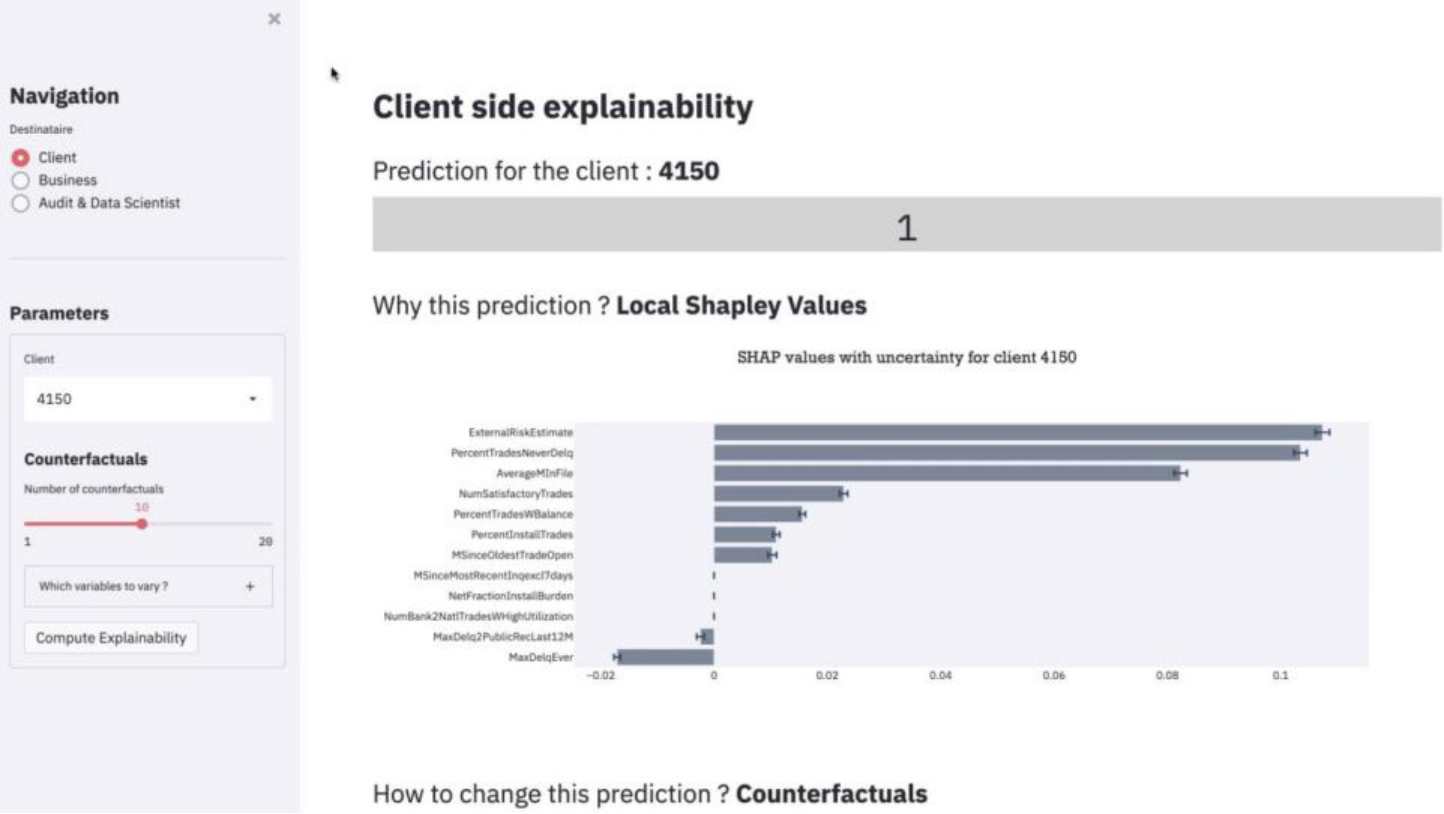
- Tentatives d'inférer la **classe générale** du modèle (linéaire, arbre de décision, …) voire ses paramètres

- **Algorithme d'apprentissage** très difficile à détecter
  - Approches heuristiques
  - Boîte à outils
  - Hacks

- **Détails algorithmiques** (hyperparamètres) encore plus élusifs

Travaux présents et futurs : toujours en XAI ...

Démonstrateur Tech Sprint



Interactions humain/machine

Étude sur les facteurs humains dans la réception d'une explication

Expérimentation sur l'explicabilité des *robo advisors* en assurance-vie

Travaux présents et futurs : autres sujets IA

Audit

Évaluation

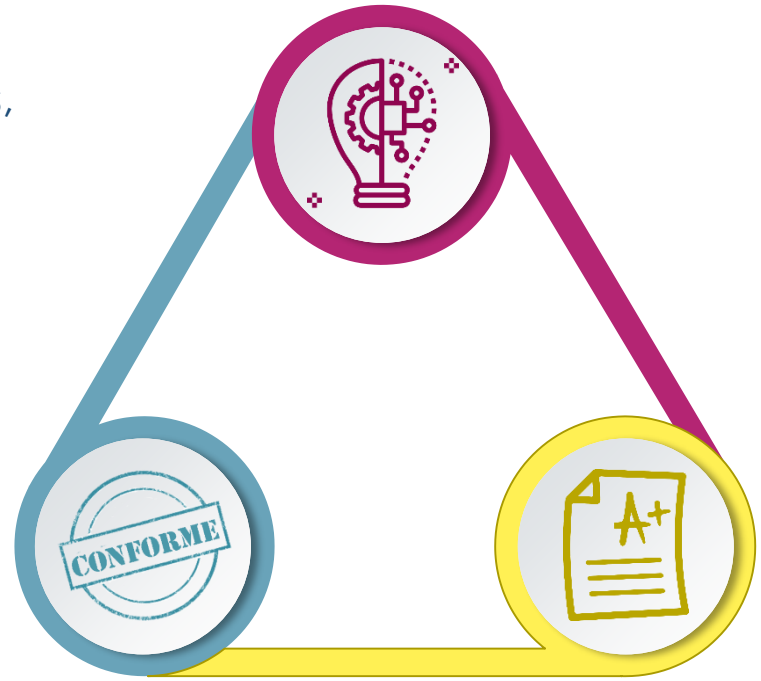
Homologation / certification

Le Grand Défi :
certifier, sécuriser et fiabiliser les systèmes fondés sur l'IA

Technological pillar :
confiance.ai

Software and methodology to support the design of trusted AI products and services

… industry strongly involved in programs, especially AI Manifesto members



… Cooperation with French basic research Initiatives, such as Aniti or DataIA, and academic research

Norms pillar

Norm, standard and regulation environment toward certification

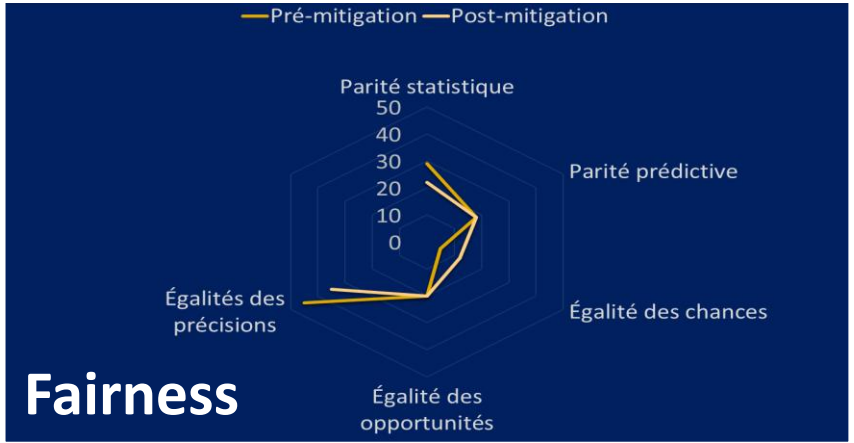
CEN CENELEC ETSI
European Standards Organizations

Applications conformity assessment Pillar

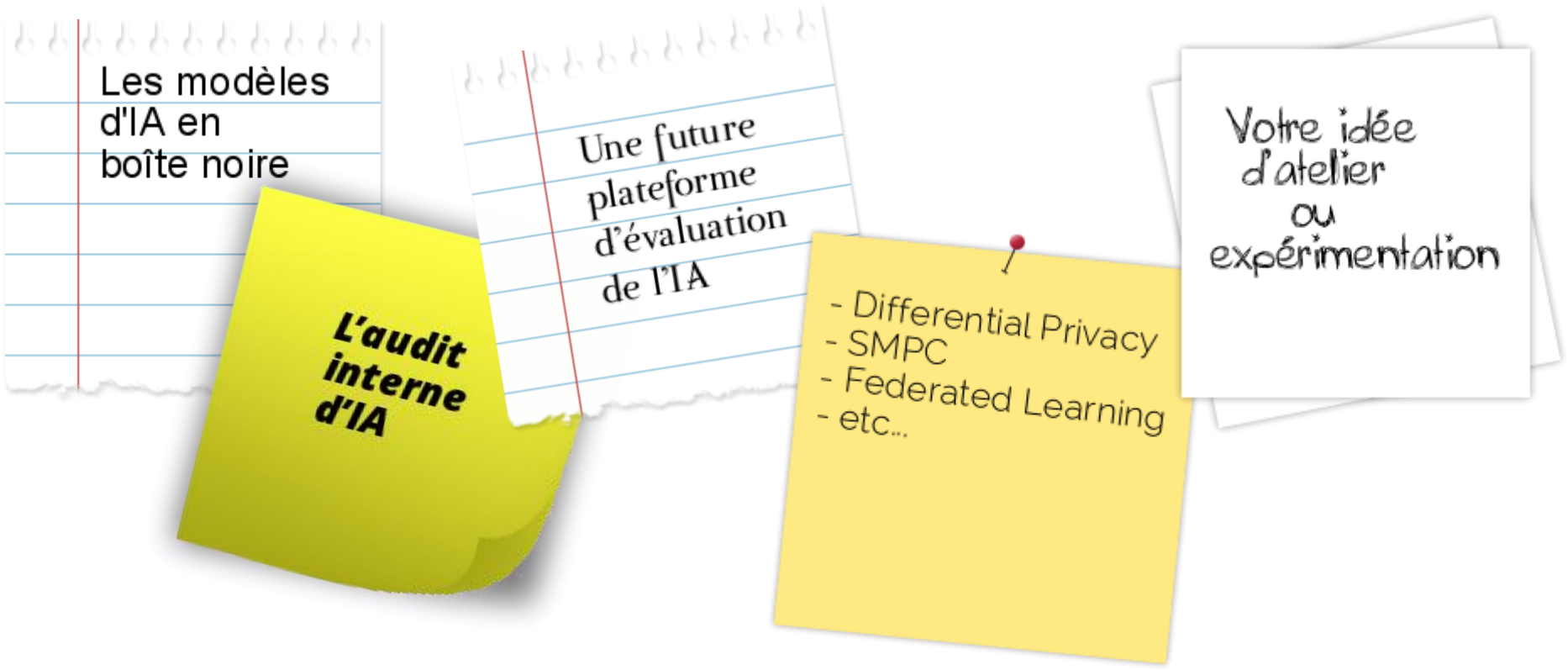
Improve evaluation of use cases

Toward a global strategy with coordinated programs and public/private funding

Travaux présents et futurs : sujets transverses



Pour échanger avec nous sur…



fintech-innovation@acpr.banque-france.fr

techsprint2021@acpr.banque-france.fr

Expérimentation sur l'explicabilité des robo-advisors en assurance-vie

Astrid Bertrand, doctorante à Télécom Paris
et l'ACPR

FORUM FINTECH 2021
Atelier IA

Les Robo-Advisors en assurance-vie

Outil algorithmique permettant de proposer un contrat d'assurance-vie avec une allocation d'actifs prenant en compte le profil et les besoins du client

1. Recueil d'informations client (objectifs d'investissement, connaissances et expérience financière, situation patrimoniale, appétence au risque) et établissement du profil de risque

2. Allocation d'actifs au sein d'un contrat d'assurance-vie

Les Robo-Advisors en assurance-vie: un cas d'IA à expliquer

Des besoins d'explicabilité envers le souscripteur

➔ Expliquer pour donner confiance

➔ Expliquer pour permettre une prise de décision éclairée

➔ Expliquer pour contrôler la pertinence du système

➔ …

- Différents objectifs d'explicabilité extraits de Liao et. al. 2020

Explications*

Un devoir d'information et de conseil (L.522-5 CdA)

➔ Formaliser les raisons justifiant le caractère approprié du contrat proposé en fonction des exigences et besoins exprimés

➔ Fournir des informations objectives sur le produit d'assurance clair, exactes et non trompeuses

Justifications*

* au sens du vocabulaire utilisé dans la littérature en explicabilité du Machine Learning (Cotton and Biran, 2017)

Des Robo-Advisors Explicables?

- Quel est l'impact des explications des robo-advisors sur le choix des souscripteurs?
- Permettent-elles une prise de décision plus "éclairée"?
- Comment différents types d'explication influencent-ils la prise de décision des souscripteurs? - Nunes & Janach, 2017
- Quels sont les besoins en explicabilité pour différents types de profils clients (connaissances et expérience financières variées, objectifs différents…)?

Quelques desiderata d'une « bonne explication »
(rappel du Tech Sprint)

Être pertinente pour un destinataire donné. Répondre aux questions contrastives de l'utilisateur

Être vraie « Rien que la vérité ». Nécessaire mais non suffisant

Être complète « Toute la vérité »

Être simple « Ne pas être écrasante »

(1) être vraie ; (2) être complète ; mais (3) ne pas être écrasante

- Kulesza, 2013

Une experimentation pour tester l'explicabilité des robo-advisors en assurance vie

1 **Concevoir** des « parcours » d'explicabilité

Pertinents – Adapter les explications au profil du client. Plusieurs parcours pour plusieurs profils.

Vrais – Mesurer la fidélité des explications et afficher l'incertitude si besoin

Complets – Contrainte de diversité au sein d'une banque de questions d'explicabilité (Liao et al., 2020).

Simples - Mesurer le nombre de morceaux cognitifs, séquencer l'information.

2 **Mesurer** l'impact des explications sur la prise de décision des souscripteurs

- ✓ Caractéristiques du choix selon les explications (risque, performance, complexité du contrat)
- ✓ Confiance de l'utilisateur dans son choix et satisfaction
- ✓ Modèle mental de l'utilisateur du robo-advisor (capacité à simuler, à comprendre les facteurs décisifs)
- ✓ Compréhension du domaine financier

Références

- Cotton, C. & Biran, Or. (2017), *Explanation and Justification in Machine Learning: A Survey.*
- Liao, Q. V., Gruen, D., & Miller, S. (2020), *Questioning the AI: Informing Design Practices for Explainable AI User Experiences.* Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,
- Nunes, I., & Jannach, D. (2017). *A systematic review and taxonomy of explanations in decision support and recommender systems.* User Modeling and User-Adapted Interaction.

Merci !

AI Act : a proposed regulation for AI products and Services

EUROPEAN
COMMISSION

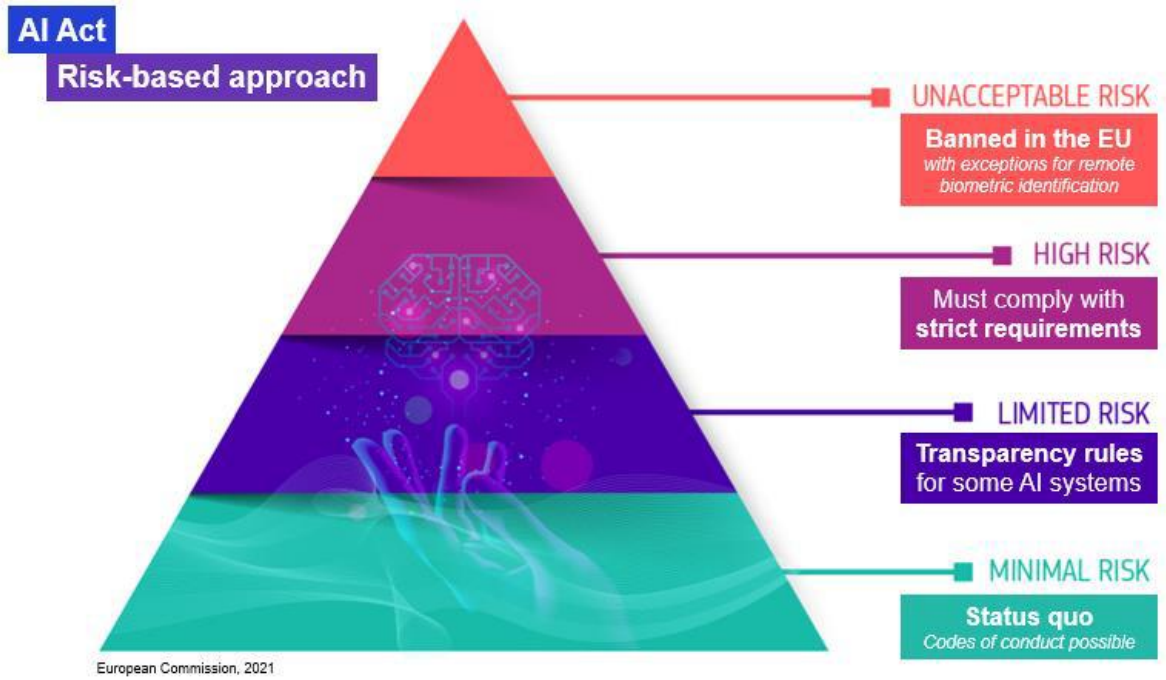
Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

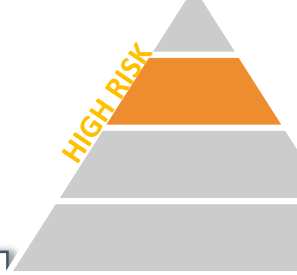
LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS



Against this political context, the Commission puts forward the proposed regulatory framework on Artificial Intelligence with the following **specific objectives**:

- ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values;
- ensure legal certainty to facilitate investment and innovation in AI;
- enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;
- facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.

Overview: obligations of operators of high-risk AI (Title III, Chapter 3)



Provider obligations

- Establish and Implement **quality management** system in its organisation
- Draw-up and keep up to date **technical documentation**
- Undergo **conformity assessment** and potentially re-assessment of the system (in case of significant modifications)
- **Register** standalone AI system in EU database (listed in Annex III)
- Sign declaration of conformity and affix **CE marking**
- Conduct **post-market monitoring**
- **Report serious incidents &malfunctioning** leading to breaches to fundamental rights
- **Collaborate** with market surveillance authorities

User obligations

- Operate high-risk AI system in accordance with **instructions of use**
- Ensure **human oversight & monitor** operation for possible risks
- Keep **automatically generated logs**
- **Inform any serious incident & malfunctioning** to the provider or distributor
- **Existing legal obligations** continue to apply (e.g. under GDPR)

Examples of applications using AI that required trustworthiness