



ILB DataLab

Une équipe spécialisée en science des données au sein du Réseau Louis Bachelier

2019

Mission

« Accompagner la transformation numérique (Data + Datascience) de l'industrie banque, finance, assurance & gestion d'actifs »



En quelques chiffres

5

ans d'Expérience

15

Data Scientists

25+

Projets réalisés

70 %

Industrialisés

20 %

En pilote

10+

Thèmes de recherche appliquée

39

Partenariats académiques

60 %

Crédit Impôt Recherche
Mécénat



Cellule de recherche appliquée

- Une équipe d'ingénieurs de recherche :
 - **Rodée aux problématiques de l'industrie financière**
 - Habituee à travailler avec des **entreprises** et des **académiques**
 - **Reconnue** pour son savoir-faire en :
 - Traitement de données et analyses statistiques
 - **Machine learning** et IA
 - Modélisation descriptive, prédictive et prescriptive
 - Proposant une approche complète de la collecte de données à **l'industrialisation**
- 3 règles de fonctionnement :
 - **Rigueur** académique
 - **Sécurité des données** : protégées par des protocoles d'accès stricts
 - **Maîtrise de la communication** : confidentialité absolue du projet ou mise en œuvre d'un plan de communication, y compris dans les publications du groupe Louis Bachelier

Une rigueur académique pour des résultats opérationnels

- Les solutions choisies mettent en avant **l'explicabilité** et **l'interprétabilité** des résultats pour répondre aux besoins de l'industrie et aux exigences des régulateurs
- Une **caution scientifique** de haut niveau avec un accès à près de **500 chercheurs de premier plan et de renommée internationale**, pour valider la pertinence scientifique
- Une équipe d'ingénieurs en data science expérimentée et formée aux techniques de pointe en Machine Learning sous la **direction scientifique de Jean-Michel LASRY** (professeur émérite à l'Université Paris-Dauphine)

Track Record démontré et coûts maîtrisés

- Equipe mobilisée, cohérente et avec un **track record significatif** : 70% des projets mis en production, 20% en phase de pilote
- Proposition ajustée pour un **impact maximal en capital humain** par l'association d'ingénieurs et d'étudiants-chercheurs
- Actions assimilables à un projet de recherche de l'Institut Louis Bachelier et bénéficiant d'une déduction de **60% de Crédit Impôt Recherche** ou des avantages fiscaux liés au mécénat

Interventions Adaptées aux Besoins du Partenaire

Conduire des programmes de recherche appliquée

Apporter des expertises spécifiques

Co-Construire des solutions jusqu'à la production

Structurer le département d'analyse des données

Identifier des solutions aux enjeux business

Les références



Île de France



NORD EST



BRIE PICARDIE
Banque & Assurance



CIB



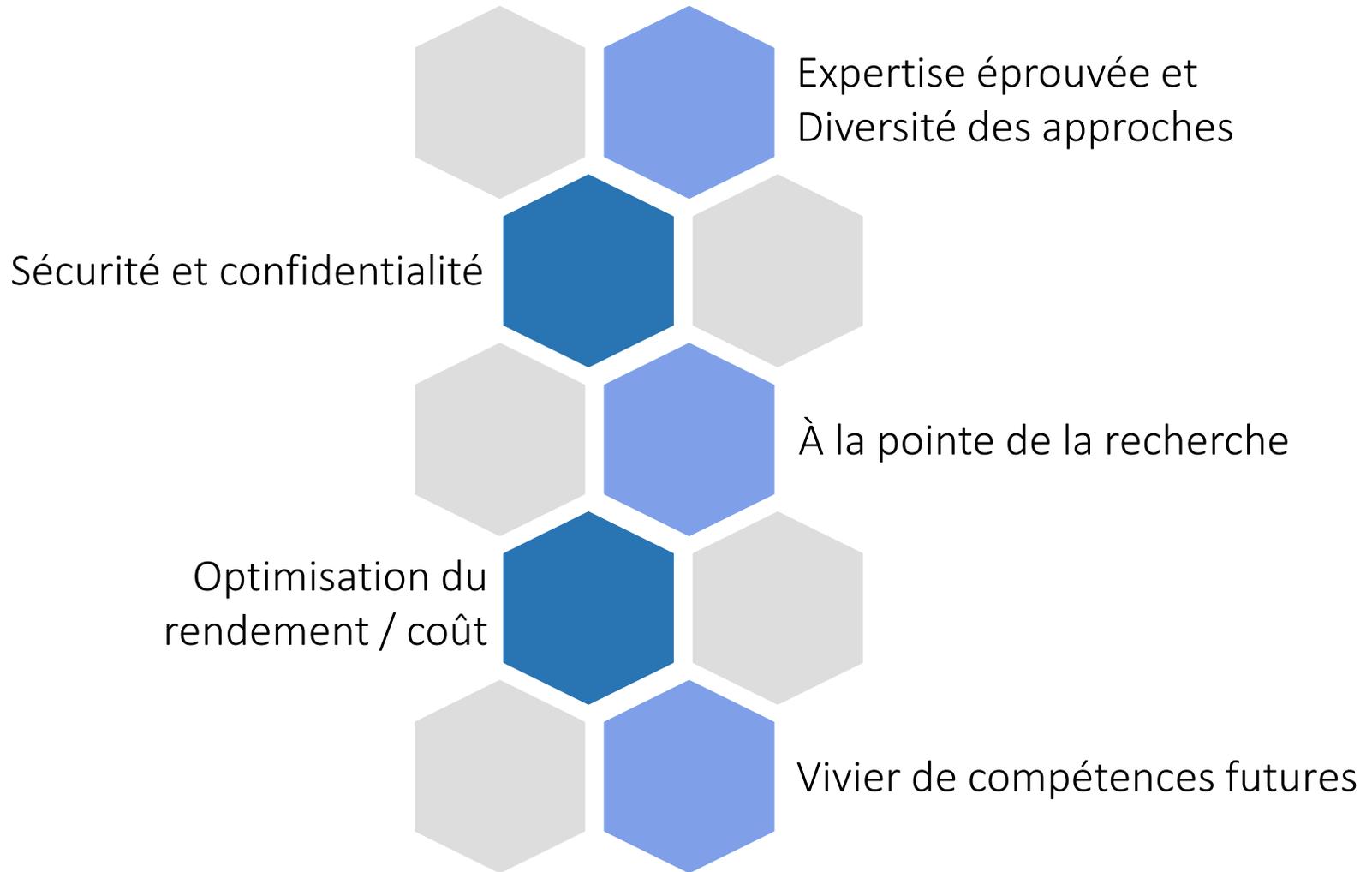
GENERALI



expert en crédit immobilier



L'ILB DataLab : une équipe complémentaire pour réussir votre Transition Numérique



L'Equipe en 2019



Jean-Michel BEACCO
Directeur Général ILB



Jean-Michel LASRY
Directeur Scientifique



Louis BOULANGER
Responsable DataLab



Driss LAMRANI
Economie & Finance
Deep Learning, NLP & Graphes



Harold OLLIVIER
Conseil
Théorie des Graphes



Krim ZIANE
Assurances
Machine Learning, Actuariat



Marine NEYRET
Finance de Marché
Modélisation Quantitative



Yuting CHEN
Banque de détail
Machine Learning



Enzo DALBY
Banque de détail
Numérique - Infrastructure



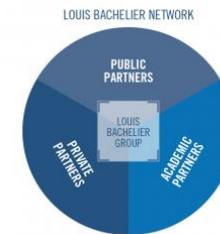
Melchior SAVIGNEUX
Conseil
NLP

Institut Louis Bachelier : un groupe et un réseau



Groupe Louis Bachelier : Moteur des activités du Réseau Louis Bachelier

Le groupe Louis Bachelier a pour objet de financer, développer et promouvoir la recherche d'excellence en Economie et Finance. Le groupe Louis Bachelier est constitué de l'Institut Louis Bachelier (ILB) et de deux fondations de recherche partenariale : « L'institut Europlace de finance » (EIF) et « La fondation du Risque » (FDR).



Réseau Louis Bachelier : Le réseau partenarial en Economie et Finance

Le Réseau Louis Bachelier a été créé pour stimuler les échanges de savoir et les collaborations transversales dans le but de favoriser l'émergence d'une économie et d'une finance durables. Le réseau mobilise des chercheurs, des entreprises privées et des partenaires publics comprenant « la Direction Générale du Trésor » et « la Caisse des Dépôts » qui ont créé l'Institut Louis Bachelier en 2008.

Ensemble, nous trouvons des réponses pour un monde en transition !



Contacts

Jean-Michel BEACCO

Directeur Général Institut Europlace de Finance

jmbeacco@institutlouisbachelier.org

+33 1 44 27 17 09

Louis BOULANGER

Responsable ILB DataLab

Louis.Boulanger@institutlouisbachelier.org

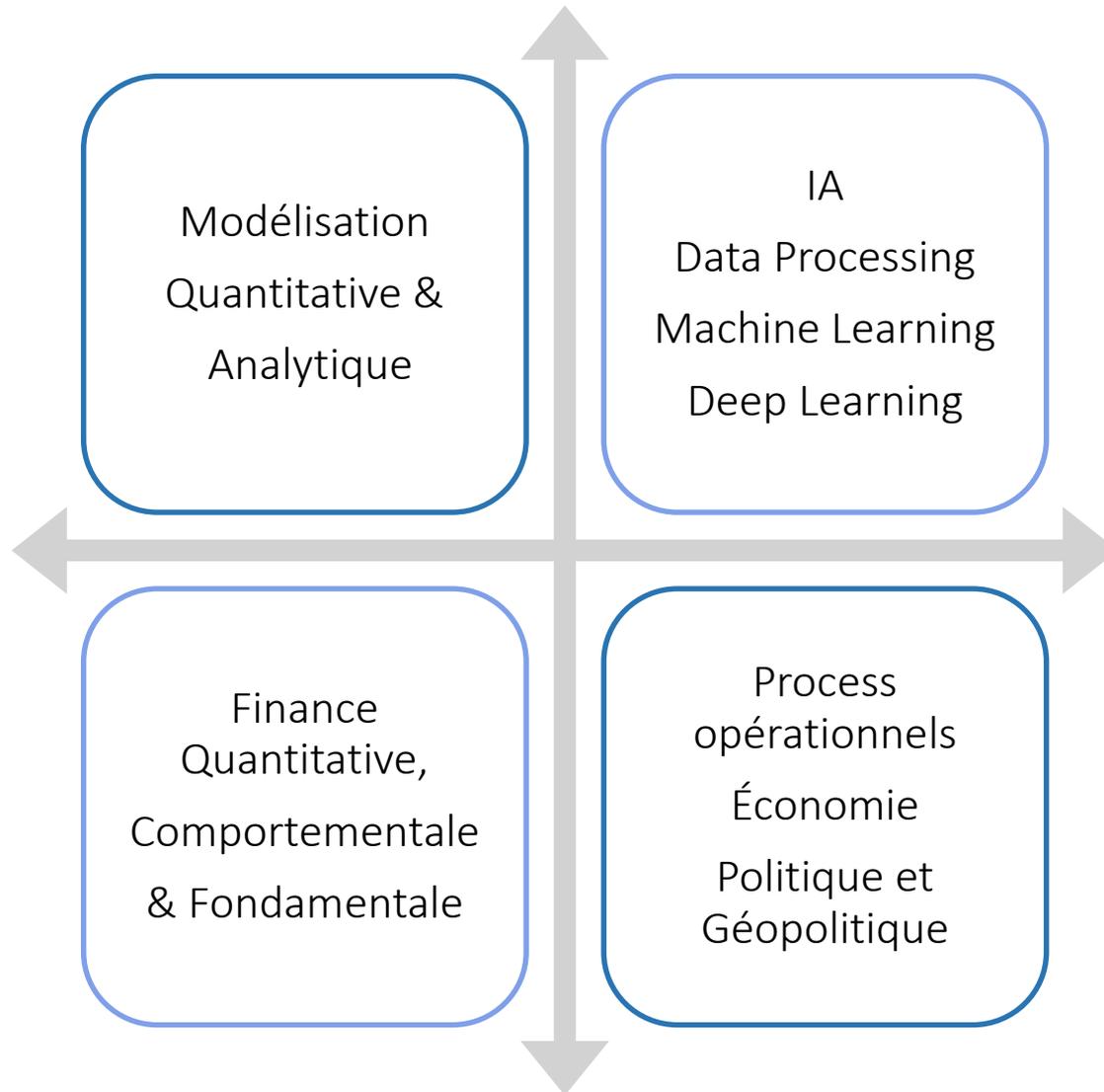
+33 6 25 53 18 57

① Projets réalisés 2014 - 2019

Typologie des projets

- Identification d'anomalies
- Collecte de données par Scraping
- Outils de visualisation
- Optimisation de la relation client
- Modélisation économétrique et financière
- Etude des comportements à partir des parcours web
- Evaluation de marché et de parts de marchés
- Prédiction des risques
- Analyse sémantique
- Etude des Graphes

Compétences de l'Equipe



Projets Réalisés (1)

- Prédiction d'impayés de prêts
- Prédiction de la performance des actions CAC-40 & DAX-30
- Prédiction de l'utilisation de matériaux médicaux
- Modèle économétrique de prévision du chiffre d'affaires
- Prédiction des revenus fiscaux des clients bancaires
- Prédiction de l'attrition consécutive aux mandats Macron
- Prédiction de l'attrition des clients bancaires
- Modélisation des parcours professionnels
- Amélioration du ROI des efforts commerciaux
- Création d'un score des places financières

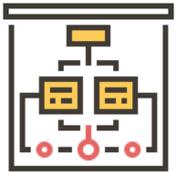
Projets Réalisés (2)

- Visualisation des corrélations des matières premières
- Évaluation taille d'un marché et taux de pénétration
- Prédiction de l'intérêt des clients à partir du parcours Web
- Optimisation du remplissage des distributeurs de billets
- Réduction des insatisfactions clients dues aux plafonds de : retraits, paiements et d'autorisation de découvert
- Anticipation des événements de vie et identification des nouveaux besoins clients
- Assistance de structuration d'une équipe de Data Science et assistance des missions

Prédiction d'impayés de prêts



- Créer un score de crédit pour les prêts hypothécaires particuliers



- Modèle auto-apprenant d'arbres de décision
- Variables : évènements de crédits, volume d'épargne, flux bancaires, indicateurs sociodémographiques et équipement en produits bancaires
- Score : probabilité d'impayés à horizon 6 mois, réactualisée

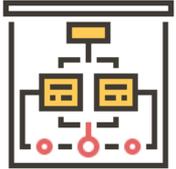


- Score 2x plus efficace que les indicateurs existants
- Score 3,2x plus efficace que la note de crédit réglementaire « Baloise »
- Focus sur les dossiers nécessitant une gestion préventive des défauts
- Cartographie de la souscription des agences / régions
- Signaux avant-coureurs de la dégradation du Score
- Identification de leviers de gestion des dossiers « à risque »

Prédiction de la performance des actions CAC-40 & DAX-30



- Anticiper la performance quotidienne close to close à 1 jour



- Modèle auto-apprenant d'arbre de décision
- Variables : indicateurs d'analyse technique (incl. moyennes mobiles, oscillateurs stochastiques, volatilité réalisée)

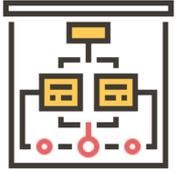


- 53% de signaux corrects
- Identification des baisses > 10 %
- Sharpe ratio : 0.8 à 1.0, incl. coûts des transactions

Prédiction de l'utilisation de matériaux médicaux



- Identifier les causes d'arrêt de l'utilisation de ventilateurs contre l'apnée du sommeil



- Modèle d'apprentissage supervisé
- Variables : données collectées par les matériaux, dossiers médicaux, description des ventilateurs, annotations d'utilisation par les patients

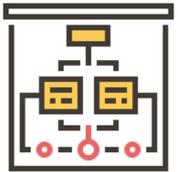


- Identification des raisons d'arrêt du traitement
- 1/3 des matériaux constituaient un défaut de calibrage gênant le patient
- Mise en place de politique correctrice de calibrage systématique

Modèle économétrique de prévision du chiffre d'affaires



- Réduire l'erreur de prévision de revenus des émissions d'actions sur le marché secondaire



- Modèle économétrique d'explication d'émissions sur le marché secondaire
- Test de la stabilité des prédictions
- Prédiction de revenu sur base d'analyses multi-scénarios macro-économiques

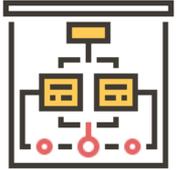


- - 45 % de l'erreur de prédiction (hors opérations exceptionnelles)

Prédiction des revenus fiscaux des clients bancaires



- Réduire l'erreur entre les revenus fiscaux des clients bancaires et les revenus collectés dans les systèmes d'information



- Modèle d'arbre de décision avec variables incomplètes
- Variables : données sociodémographiques, volume d'équipement, flux bancaires et volume d'épargne (flux et stock)
- Partenariat avec un chercheur de l'INRIA notamment pour l'optimisation des calculs : mémoire et temps d'exécution

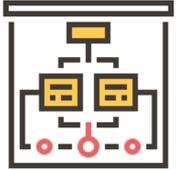


- Distribution de revenus fiscaux à partir d'informations incomplètes
- Identification des personnes à hauts revenus avec une marge d'erreur tolérée
- Redéfinition des segmentations clients
- Réduction de l'erreur de prédiction des revenus de 50 %
- Déploiement en production sur le réseau bancaire après une phase test

Mesure des fonds propres des bancassureurs



- Étude commanditée par la Banque Centrale Européenne
- Construire des approches alternatives de fonds propres et les comparer aux exigences réglementaires



- Modèle de mesure des besoins Solvabilité 2 et Bâle 4 par catégorie d'activité
- Consolidation des besoins prenant en compte la diversification
- Sensibilité de consommation du capital par source

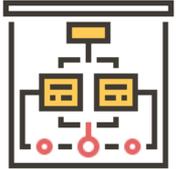


- Option Danoise (retenue par la BCE) : meilleure approche de mesure des fonds propres

Prédiction de l'attrition consécutive aux mandats Macron



- Prédire les comptes bancaires en risque d'arbitrage « mandats Macron »
- Anticiper la clôture des comptes



- Modèle d'arbre de décision
- Variables : flux bancaires, évolution de la taille d'épargne, utilisation des moyens de paiements

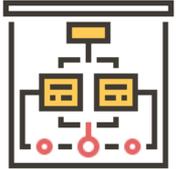


- Construction de signaux mensuels de risque de clôture dans 3 mois
- Politique marketing pour les comptes à risque

Prédiction de l'attrition des clients bancaires



- Construire des signaux prédictifs du risque de fermeture de compte



- Modèle d'arbre de décision
- Variables : flux bancaires, évolution de la taille d'épargne, utilisation des moyens de paiements

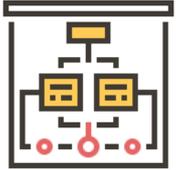


- Construction de signaux mensuels de risque de clôture dans les 6 mois
- Réactualisation de l'apprentissage et suivi de la performance du modèle
- Externalité : identification des « événements de vie » à partir de l'analyse réalisée

Modélisation des parcours professionnels



- Modéliser et comprendre les transitions professionnelles



- Préparer, nettoyer et enrichir les données de 1,5 millions de CV
- Construire des matrices de transitions des parcours professionnels
- Identifier les co-occurrences des compétences aux transitions

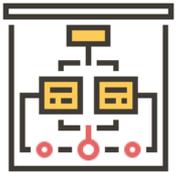


- Production d'une taxonomie des compétences informatiques
- Visualisation en graphiques de Sankey des transitions professionnelles
- Évaluation de la période moyenne de maintien en poste
- Recommandation de parcours professionnels
- Recommandation de formations en informatique et langues
- Identification des facteurs de tension régionale de l'offre du marché de travail

Amélioration du ROI des efforts commerciaux



- Identifier les leviers d'amélioration du ROI
- Recommander des évolutions stratégiques et opérationnelles



- Modèle de rentabilité analytique du démarchage commercial
- Identification des facteurs saisonniers, des opportunités et de leurs qualités
- Priorisation des démarchages par typologie de client et de demande

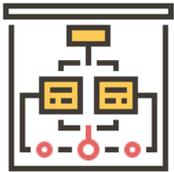


- Baisse de 1/3 des appels sans baisse de chiffre d'affaires
- Flexibilisation de la gestion de l'équipe commerciale

Création d'un score des places financières



- Challenger le score « Global Financial Centers Index » des places financières



- Organisation d'un challenge étudiant pour l'évaluation des places financières avec le soutien du Ministère des Finances et sous le patronage d'André Lévy-Lang

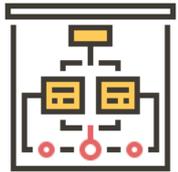


- Synthèse de recommandations sur la définition de l'attractivité, la mesure, la construction d'un score et les écueils à éviter

Visualisation des corrélations des matières premières



- Visualiser les corrélations des marchés financiers



- Méthode d'analyse de graph et de visualisation développées en collaboration avec l'Université Paris-Dauphine

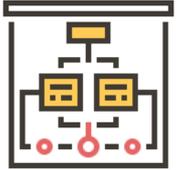


- Outils de visualisation temporelle de la corrélation entre marchés
- Construction d'indicateurs de propagation des chocs financiers

Évaluation de la taille d'un marché et du taux de pénétration



- Évaluer la taille locale du marché du crédit consommation
- Mesurer le taux de pénétration locale des agences bancaires



- Collecter des données open-data : INSEE
- Construire des algorithmes d'estimation de la volumétrie des crédits octroyés

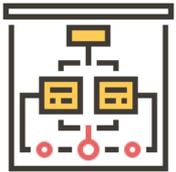


- Rationalisation des objectifs des agences bancaires pour le crédit consommation

Prédiction de l'intérêt des clients à partir du parcours Web



- Évaluer l'efficacité des actions marketing post parcours site internet



- Mesure du taux de souscription post consultation des pages d'information
- Évaluation des taux de transformation des communications par SMS, E-mail et contact téléphonique
- Redéfinition du ciblage des clients des centres d'appels

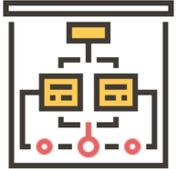


- Identification de parcours clients inefficaces
- Contact par SMS inefficace et coûteux
- Contact téléphonique plus efficace que l'envoi d'e-mails
- Identification d'un groupe de client (30 % de la population) dont le potentiel est 40 % plus élevé
- Réduction de coût d'action marketing et amélioration du taux de transformation

Optimisation du remplissage des distributeurs de billets



- Optimiser la logistique du remplissage de billets sous contraintes de risque de vol, coûts d'assurance et réapprovisionnements exceptionnels



- Mesure de la distribution du besoin monétaire entre deux réapprovisionnements programmés
- Optimisation dynamique sous-contraintes du besoin monétaire

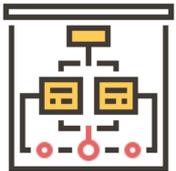


- Recette des tests en cours pour une mise en production fin 2019

Réduction des insatisfactions clients dues aux plafonds de retraits, de paiements et d'autorisation de découvert (1)



- Évaluer les plafonds des moyens de paiement de découvert autorisé en entrée en relation et en cours de vie
- Réduire les insatisfactions clients :
 - 1/4 des clients ont expérimenté un refus de retrait
 - 1/2 des clients ont expérimenté un refus de paiement
 - 1/2 des clients ont expérimenté un incident de découvert



- Optimisation des plafonds pour réduire les insatisfactions
- Construction de contraintes de revenus en lien avec la gestion des risques
- Focalisation sur les clients avec une relation commerciale supérieure à 12 mois

Réduction des insatisfactions clients dues aux plafonds de retraits, de paiements et d'autorisation de découvert (2)

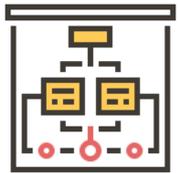


- Plafonds de retrait :
 - - 25 à - 50 % des insatisfactions liées aux retraits (en fonction des revenus)
 - + 43 % des plafonds de retrait
- Plafonds de paiements :
 - - 70 % des insatisfactions liées aux paiements (en fonction des revenus)
 - - 57 % des plafonds en volume
- Autorisation de découvert :
 - - 36 % à - 54 % des insatisfactions totales liées aux incidents de découvert (en fonction des revenus)
 - + 15 % des autorisations de découvert en volume total
 - - 20mn / jour à - 30mn / jour du temps des conseillers
- Prochaine étape : généralisation aux plafonds établis à l'entrée en relation

Anticipation des événements de vie et identification des nouveaux besoins clients



- Identifier les événements de vie à partir de l'analyse des paiements
- Prédire le besoin client pour le crédit consommation



- Modèle d'identification de changement de structure des dépenses
- Filtrage des clients avec des fréquences de paiements supérieurs à une transaction par jour
- Allocation des dépenses par moyen de paiement aux catégories de terminaux : 900 catégories « Merchant Category Code » regroupées en 250 catégories (sous contrainte de RGPD et CNIL)

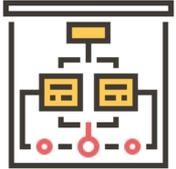


- Identification des régimes stationnaires de dépenses
- Rapprochement des résultats du modèles aux événements de vie identifiés (i.e. naissances)
- En cours : test auprès des conseillers pour revue et amélioration des modèles

Assistance de structuration d'une équipe de Data Science et assistance des missions



- Identifier les fraudes dans les marchés de capitaux
- Transférer les compétences vers l'équipe interne

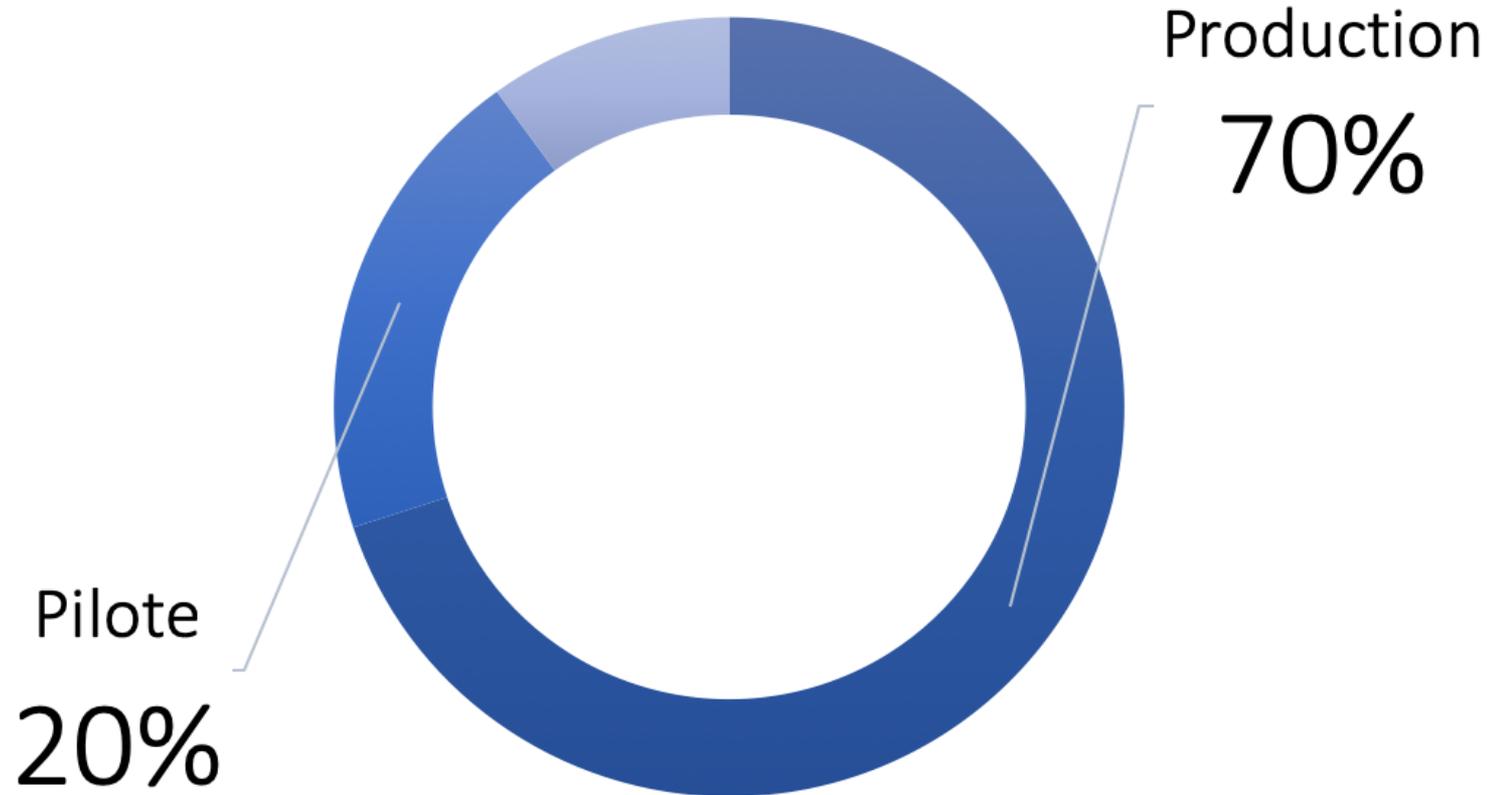


- Construction de modèles d'apprentissage profonds pour identifier les opérations frauduleuses



- 15 % de baisse des fausses alertes

Industrialisation des Preuve de Concepts

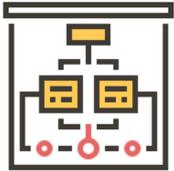


② Thèmes de recherche appliquée

Mesure de la toxicité sur les réseaux sociaux



- Mesurer la toxicité des commentaires en langage naturel (anglais) pour faciliter l'activité des modérateurs
- Apprendre la subtilité linguistique : sarcasme, humour, insultes, menaces...
- Résoudre les problématiques du faible nombre d'exemples d'apprentissage



- Construction de 3 modèles d'analyse sémantique supervisés
- Apprentissage sur une base de 1,8 millions de commentaires annotés

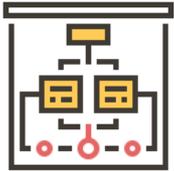


- 94 % d'efficacité du modèle
- 70 % d'efficacité du sarcasme / humour vs. toxicité

Prédiction d'anomalies à partir de données d'indicateurs IoT



- Prédire l'occurrence d'un tremblement de terre à partir des données sismologiques



- Construction de 5 modèles d'apprentissage profonds et d'arbre de décision



- 92.5 % d'efficacité de prédiction de tremblement de terre simulés par le laboratoire Los Alamos National Laboratory (U.S.)

Prédiction des risques financiers par l'analyse de la presse



- Identifier l'impact des nouvelles financières (annonces, communiqués de presse, publications) sur l'évolution des risques financiers



- Construction de base de savoir financier
- Construction de modèles d'apprentissage profonds pour identifier les entités, les relations sémantiques et la sensibilité des risques financiers aux nouvelles



- 65 articles traités par seconde (sous contrainte d'infrastructure)
- Test en cours de l'efficacité de prédiction des risques financiers

Typologie des projets

- Évaluation du risque de crédit PME / ETI pour la titrisation des créances
- Évaluation du risque assurantiel à partir des commentaires sur les réseaux sociaux
- Modélisation géologique de la production pétrolière : structure, efficacité, et identification d'anomalie
- Organisation de l'équipe Data Science d'une FinTech
- Construction d'indicateurs de cycles macro-économiques et de stabilité financière
- Construction d'indicateurs de tension sur l'offre immobilière
- Modélisation des mises en adéquation pour un site de rencontres
- Prédiction du temps d'arrivée des livraisons de denrées périssables