Square

DONNER DU FUTUR AU TALENT

DATA

LE MACHINE LEARNING EN FINANCE

VERS UNE VÉRITABLE RÉVOLUTION ?

ADRIEN CAIAZZO,
CHAÏTANYA NA CHAMPASSAK,
GAÏANE FLOCH,
HAMID SALEM.

ÉTUDE I JUILLET 2020

Square

DONNER DU FUTUR AU TALENT

DATA

LE MACHINE LEARNING EN FINANCE

VERS UNE VÉRITABLE RÉVOLUTION ?

Par Adrien CAIAZZO, Chaïtanya NA CHAMPASSAK,
Gaïane FLOCH, Hamid SALEM

SOMMAIRE

Square

DONNER DU FUTUR AU TALENT

INTRODUCTION

Les avancées technologiques de ces 20 dernières années, telles que la blockchain ou le Big Data, sont à l'origine de nombreux bouleversements humains mais aussi sociétaux. L'entrée dans un monde de plus en plus numérique a décuplé la création de données et l'exploitation de celles-ci à grande échelle. Cette exploitation s'appuie sur des théories mathématiques et statistiques dont les possibilités de mise en œuvre s'accélèrent avec les progrès en matière de calcul. Parmi les disciplines rattachées à la science de la donnée, le Machine Learning connaît une expansion rapide.
Ce focus se destine à un public non averti désireux d'en savoir plus sur les usages actuels du Machine Learning au sein des services financiers.

Encore obscur voire abstrait pour le grand public, le Machine Learning a su faire sa place au sein des plus grandes industries et fait désormais partie du paysage de nombreuses activités. Par exemple, dans le domaine de la distribution et plus particulièrement dans les grandes surfaces, les rayons y sont désormais organisés pour booster les ventes grâce à l'apprentissage du Machine Learning (emplacement des produits, choix des couleurs, etc.). Cette technologie permet également une réelle optimisation de la gestion des stocks pour les distributeurs. C'est à ce titre que l'enseigne Carrefour s'est offerte les services de Google à partir de novembre 2018 pour optimiser les assortiments proposés ou encore la prévention des ruptures en linéaires.

La productivité ne cesse de s'améliorer et de nouvelles méthodes de travail apparaissent : les industriels anticipent d'ores et déjà les pannes des équipements afin d'augmenter la productivité d'une usine. Depuis 2013, Fujitsu utilise cette technologie avec une augmentation de la productivité ayant atteint 20%, tout en étant couplée à une réduction de 10% des coûts globaux liés à la maintenance.

Le secteur financier n'est pas en reste, même si le virage du Machine Learning et de ses variantes n'a été adopté que tardivement ou partiellement dans la plupart des établissements. En 2017, seuls 7% des acteurs financiers français avaient adopté le déploiement d'outils robotiques simples (contre 40% dans le monde).

Le Machine Learning, nouvelle révolution technologique aux possibilités infinies, est souvent mis en avant comme une évolution à l'origine de ce progrès. Autrefois asservi à l'Homme, l'ordinateur devient autonome, capable d'apprendre de lui-même, multipliant ainsi le champ des possibles.

Souvent avancé comme technique, coûteux et difficile à mettre en place, le Machine Learning est-il une véritable révolution ? Dans quelle mesure bouleverse-t-il les modes de travail au sein des entreprises ? Quelles sont les utilisations possibles pour les clients et pour les entreprises ?

L'enjeu de ce document est de comprendre, aujourd'hui, le fonctionnement du Machine Learning, son intérêt mais avant tout d'apporter un éclairage à la question suivante : « Le Machine Learning en finance, une révolution, mais pour quels usages ? ».


imgIX
imgIX
ORTRONICS
ORTRONICS

1.

INTRODUCTION AU MACHINE LEARNING

1.1 L'UNIVERS DU MACHINE LEARNING

1.1.1 Histoire et définition du Machine Learning

Le Machine Learning, aussi appelé apprentissage automatique, se définit comme une technologie d'apprentissage artificiel, elle-même dérivée de l'intelligence artificielle.

Si l'ordinateur effectue habituellement un programme défini procurant des résultats, l'apprentissage automatique va ici conférer au programme la capacité d'apprendre de ses actions et des résultats obtenus.

L'ordinateur produit des modèles et les fait évoluer au fur et à mesure qu'il intègre de nouvelles données.

Les algorithmes de Machine Learning ont la particularité d'utiliser d'importantes quantités de données, ce qui explique notamment leur utilisation croissante (nous n'avons jamais eu accès à d'aussi importants jeux de données qu'aujourd'hui). Ces algorithmes sont d'autant plus efficaces que les capacités de stockage de données et les puissances de nos machines augmentent.

Le Machine Learning n'est pas une technologie récente. La terminologie apparaît pour la première fois en 1959 : c'est Arthur Samuel, un informaticien américain, qui a en effet été le premier à faire usage de cette expression après avoir travaillé plusieurs années sur un programme de sa création pour IBM. Le programme en question jouait au jeu de dames et surtout, apprenait en jouant. Une révolution commençait alors.

Les capacités technologiques évoluant, le potentiel d'apprentissage des programmes fit de même. En témoigne par exemple, la victoire de Deep Blue, un autre projet d'IBM, qui fut le premier à vaincre Garry Kasparov alors champion du monde d'échecs en 1997.

Les exemples se multiplient ensuite avec en 2011, la victoire par traitement de langage de Watson

au Jeopardy, un célèbre jeu de connaissances culturelles américain. En 2015, par la victoire d'AlphaGo, Google démontre de nouveau son savoir-faire avec la création d'un très puissant programme ayant gagné au jeu de Go contre l'un des meilleurs joueurs mondiaux. Plus récemment, AlphaStar, toujours de Google, a battu les meilleurs joueurs mondiaux au jeu vidéo de stratégie en temps réel StarCraft II.

Le Machine Learning donne ainsi la possibilité de produire des règles ou des modèles capables d'expliquer les données, d'en prédire de nouvelles (predictive analytics), voire de prendre des décisions, comme le montrent les exemples des véhicules autonomes et des diagnostics de cancer.

Si aujourd'hui, la reconnaissance d'images fait partie intégrante de nos vies, via nos smartphones ou encore certains sites web tels que Youtube (capables de détecter le contenu des vidéos postées), les évolutions devraient devenir de plus en plus significatives. En effet, ces éléments, combinés avec d'autres types de capteurs préparent par exemple l'arrivée de véhicules 100% autonomes dans les années à venir (certains étant déjà en circulation).

La combinaison de cette technologie et de la reconnaissance d'image confère ainsi au véhicule la faculté d'identifier les obstacles devant lui et d'adapter lui-même sa réaction. D'une autre manière, les scanners utilisant le Machine Learning acquièrent la compétence de diagnostiquer un cancer chez un patient. L'algorithme va, pour cela, étudier de nombreuses images afin d'identifier des critères de détection.

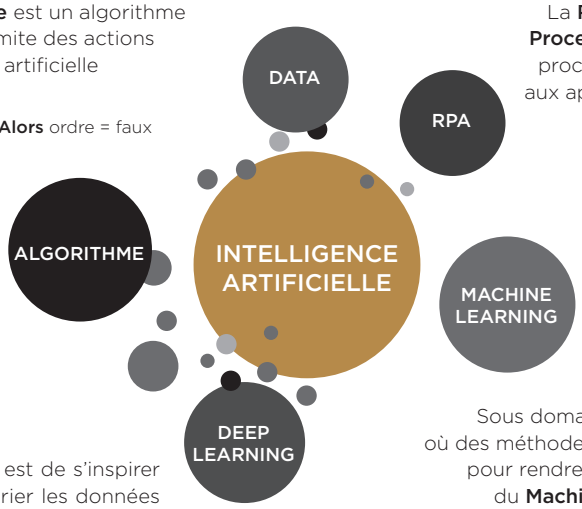
En soit, les possibilités sont aussi vastes que les idées que nous aurons.

Figure 1. Les différentes formes d'intelligence artificielle

Une **intelligence artificielle** est un algorithme plus ou moins évolué qui imite des actions humaines, une intelligence artificielle n'apprend pas.

Exemple: **Si** montant > 10000 **Alors** ordre = faux

La **RPA (Robotic Automatisation Process)** permet d'automatiser des processus métier. Elle se connecte aux applications, gère des données, exécute des tâches et se déconnecte grâce à l'IA.



L'enjeu du **Deep Learning** est de s'inspirer du cerveau humain pour trier les données utiles à l'algorithme sans aide humaine.

Sous domaine de l'intelligence artificielle où des méthodes statistiques sont appliquées pour rendre l'algorithme intelligent, le but du **Machine Learning** est d'approximer les données afin d'en tirer une analyse.

LE « DEEP LEARNING : UNE SIMPLE ÉVOLUTION ? »

Le Machine Learning est donc une puissante méthode d'analyse statistique de flux de données. Il en reste néanmoins un domaine nécessitant encore une intervention humaine au préalable.

En effet, le travail de la qualité de données est au cœur de l'efficacité intrinsèque du procédé de Machine Learning qui ne peut exploiter de données que si celles-ci sont lisibles pour l'algorithme.

Ce travail, d'une grande lourdeur et chronophage, peut aller à l'encontre même de l'utilité du Machine Learning. C'est là qu'intervient le Deep Learning, un sous-domaine du Machine Learning reposant sur un système d'analyse neuronal.

Cette technologie va plus loin que le Machine Learning dans le sens où elle comprend et extrait elle-même les données importantes.

Outre la puissance et les capacités de calcul, le Deep Learning se caractérise avant tout par la possibilité pour l'ordinateur d'analyser de très importantes quantités de données sans qu'elles ne soient organisées.

Cette typologie d'architecture permet notamment de procéder à de la reconnaissance faciale, de composer des titres musicaux ou encore de conduire un véhicule autonome.

A l'instar du processus de Machine Learning, la performance du Deep Learning croit au fur et à mesure que l'on fournit des données. L'amélioration de la pertinence des résultats croit d'autant plus vite que la quantité des données explose, expliquant l'intérêt grandissant des grandes compagnies pour l'acquisition de données clients.

1.1.2 Modèles supervisés et non supervisés

Créer un modèle de prédiction implique la formalisation de plusieurs problématiques. Que désirons-nous observer ? Que désirons-nous prédire ?

Prenons l'exemple suivant : nous disposons d'un set de photographies d'animaux (chiens, chats et tortues) et désirons construire un modèle qui reconnaît les tortues.

Deux possibilités nous sont permises :

> Nous appliquons une étiquette « tortue » sur un échantillon de photographies de tortues et utilisons un modèle algorithmique qui va se servir de cette information de base pour retrouver toutes les autres photographies présentant cet animal ;
> Nous n'appliquons aucune étiquette et utilisons un modèle qui va comprendre par lui-même la différence entre les animaux pour les classer. Il convient, dans ce cas, d'utiliser un modèle différent de celui mentionné au premier point.

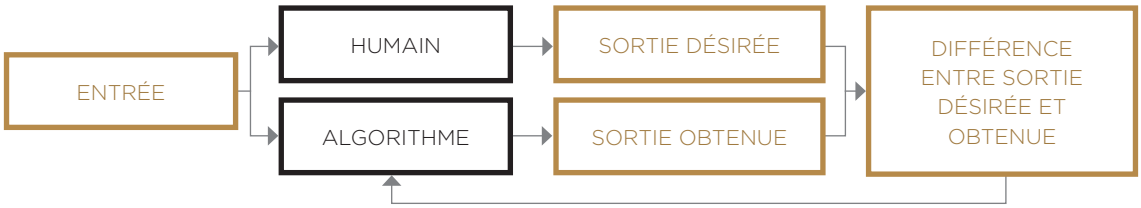
Ce principe est celui de la labélisation des données. Il est primordial dans le cadre du choix du modèle qui sera utilisé.

Deux catégories de modèles de Machine Learning s'offrent alors à nous : les modèles d'apprentissage dits supervisés (a) et les modèles d'apprentissages dits non supervisés (b).

a. Dans le cadre de l'apprentissage supervisé, un data-scientiste est présent pour définir les paramètres ou les résultats attendus par l'algorithme.

Dans le cas d'un modèle **d'apprentissage supervisé** comme la classification, ce dernier regroupe les individus en groupes homogènes définis à l'avance par l'intervenant, c'est-à-dire que l'intervenant fournit à l'algorithme des exemples avec des résultats. C'est notamment le cas pour le diagnostic des cancers. Ce modèle utilise donc des données labélisées.

Figure 2. Apprentissage supervisé



b. A contrario, **l'apprentissage non supervisé** ne nécessite pas d'intervention humaine : l'algorithme comprend par lui-même les différences ou les corrélations pour proposer le meilleur résultat, selon lui. Il va ainsi partitionner et classer les données dans des groupes homogènes, comme cela peut être le cas au travers de la méthode de « Clustering[1] ». Ce modèle ne nécessite donc pas de données labellisées.

Figure 3. Apprentissage non supervisé

De cette façon, l'apprentissage ne se fait pas à partir d'indications pré-déterminées mais à partir des fluctuations observables et observées au sein du jeu de données.

L'avantage de cette solution est qu'elle ne nécessite pas de grands volumes de données étiquetées et est moins consommatrice en ressources humaines. L'apprentissage non-supervisé est notamment utilisé pour mettre au point des systèmes de recommandation, par exemple en conseillant un film en fonction de ceux regardés précédemment. Cependant il n'effectue aucune classification, et est donc complémentaire avec l'apprentissage supervisé.

Nous nous trouvons ainsi face à deux typologies de fonctionnement aux objectifs pouvant être différents et présentant tous deux des intérêts en termes d'automatisation et d'apprentissage.

En synthèse, le Machine Learning a pour but de faire ressortir des informations de manière automatique à partir d'un gros volume de données. L'algorithme utilisé aura la faculté d'assimiler des quantités de données toujours plus grandes afin de générer des résultats.

Les résultats seront d'autant plus précis et pertinents que le volume de données sera massif. En total contraste avec le raisonnement humain qui aura plus de difficultés à gérer des quantités trop volumineuses d'informations, et surtout sera incapable de détecter une tendance (un pattern) que l'algorithme trouvera.

1. *Méthode d'analyse des données visant à diviser un ensemble de données en différents « paquets » homogènes.*

1.2 LE BIG DATA AU SERVICE DU MACHINE LEARNING

1.2.1 Qu'est-ce que le Big Data ?

L'essor du Machine Learning est fortement lié à l'augmentation du volume et de la variété des données disponibles. A travers nos aptitudes grandissantes à cumuler des données de tous genres, nous avons désormais accès à des informations d'une grande richesse et potentiellement d'une grande précision.

Les outils analytiques traditionnels ne sont pas suffisamment performants pour exploiter au mieux ces quantités de données.

C'est là qu'intervient le Big Data, caractérisé par trois dimensions fondamentales :

> **LE VOLUME :** la massivité du volume de données est l'un des critères qui caractérise le mieux le Big Data, mais ce n'est pas le seul. *Exemple : l'étude d'une base clients portant sur 20 ans d'observation donnera un volume plus important que pour une semaine d'observation, ce qui peut constituer un avantage dans certaines typologies d'analyses.*

> **LA VARIÉTÉ :** il est intéressant de disposer de données variées qui apportent des informations complémentaires les unes par rapport aux autres ; *Exemple : la multitude de données (informations sur la personne, ses habitudes de consommation, son épargne, sa mobilité) permettra d'obtenir une plus grande lisibilité et de créer des liens entre elles.*

> **LA VÉLOCITÉ :** la vélocité va caractériser la fréquence à laquelle les données vont être générées et mises à jour.

Dès lors que ces trois dimensions sont réunies, nous sommes techniquement face à ce que l'on nomme Big Data, ou presque. D'autres conditions vont être nécessaires pour être en mesure de les exploiter, à commencer par la capture et le stockage de ces données afin de les centraliser et de créer un environnement permettant leur manipulation.

C'est avec cet environnement (qui se présentera le plus souvent sous la forme technique de Datawarehouse[2] ou de Datalake[3]) qu'il sera possible de procéder à la recherche et au partage de ces informations.

1.2.2 Les sources de données du Big Data

Il existe de nombreuses sources de données aux caractéristiques très diverses. Dans les sujets appliqués à la finance, certaines sources vont être régulièrement utilisées de par leur utilité, à savoir :

- **Les données internes :** du fait de son activité, l'établissement collecte d'importantes quantités de données sur sa clientèle, sur leurs opérations, mais également sur leurs profils.
- **Les données externes :**
  - > Issues des marchés financiers : ces derniers sont des sources extrêmement importantes pour les établissements financiers ; les données issues des marchés financiers permettent d'en savoir plus sur les mouvements passés, mais également actuels, afin de pouvoir être corrélés avec d'autres informations.
  - > Issues des statistiques nationales : les informations sur l'économie, les pays et les populations sont également très appréciées car elles permettent d'en savoir plus sur un territoire. Souvent publiques, ces informations sont fournies par des institutions telles que l'INSEE.
  - > Issues de l'open data : Il s'agit de données libres d'accès. Elles doivent remplir certains critères : la disponibilité, la réutilisation, la distribution, et la participation universelle (aucune discrimination contre les personnes qui peuvent utiliser les données, par exemple pas de restriction non commerciale).

2. *Base de données utilisée pour collecter, ordonner, et stocker des informations provenant de bases de données opérationnelles. L'ensemble des données stockées servent notamment à l'aide à la décision en entreprise.*
3. *Méthode de stockage des données utilisée par le Big Data. Ces données sont gardées dans leurs formats originaux ou sont très peu transformées.*


fn
control
option
command
Q
W
A
S
D
E
X
C
V

2.

LE MACHINE LEARNING APPLIQUÉ À LA FINANCE

Les OPCVM (organismes de placement collectif en valeurs mobilières), instruments d'épargne très prisés des ménages ou des entreprises, sont un agrégat de produits financiers (actions, obligations, fonds monétaires, etc.) intégrés au sein d'un même fonds. Ce dernier est sensible aux variations de l'ensemble des valeurs qui le composent.

POUR INFORMATION : fin 2018, le marché français de la gestion d'actifs représentait 3 674 milliards d'euros (soit environ 17% du marché européen) pour plus de 11 000 fonds disponibles.

2.1 LA GESTION D'ACTIFS

2.1.1 Application du Machine Learning à l'allocation d'actifs

La difficulté pour le client va être de choisir un type de fonds pour investir et les héberger dans son produit d'épargne. Les combinaisons sont multiples et le client a besoin généralement de l'expertise de son conseiller pour l'accompagner dans ses choix. Car, bien en amont des rendements espérés par un épargnant, il existe un véritable travail d'analyse réalisé par les sociétés de gestion dotées de nombreuses compétences : les gérants de portefeuilles, les analystes quantitatifs ou les risk managers, dont les expertises servent in fine la performance servie aux clients en contrepartie de frais de gestion (en moyenne 3,47% en 2018).

Les gérants de portefeuille, aidés par les analystes quantitatifs, doivent déterminer la meilleure allocation possible pour les fonds sous gestion. L'allocation d'actifs consiste en la définition d'une stratégie d'investissement en fournissant un portefeuille modèle (soit la définition d'une stratégie de répartition entre plusieurs catégories d'actifs).

Le recours au Machine Learning prend ici tout son sens avec une suggestion personnalisée en fonction de l'aversion au risque du client, totalement automatisée et venant remplacer ou seconder le conseiller en patrimoine. Ce type de gestion de fonds pilotée permet de ne plus faire face aux aléas humains et de recourir à des algorithmes présentant une couverture large et à même de répondre aux attendus des

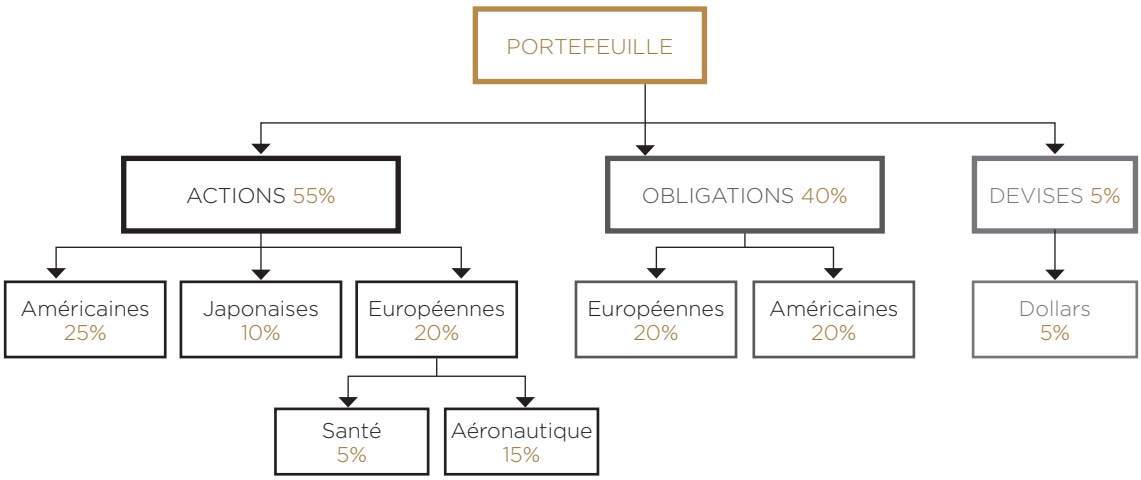
investisseurs. Ils ont par ailleurs l'avantage d'être moins coûteux qu'une gestion humaine.

La répartition des fonds ainsi que leur évolution dans le temps seront observées et travaillées par la machine qui pourra avoir, dans certains cas, un rôle totalement autonome, ou semi-autonome (de l'ordre de la suggestion et non de l'action).

L'exposition du portefeuille est calculée en fonction de la classe d'actifs, du pays ou encore par secteur d'activité.

Un portefeuille modèle pourrait ainsi avoir cette apparence :

Figure 4. Exemple de portefeuille de gestion



Une stratégie d'allocation va être généralement décomposée en plusieurs phases :

> Déterminer les classes d'actifs (exemple : 55% d'actions, 40% d'obligations et 5% de devises étrangères).

> Identifier les localisations (il est ici préférable d'avoir des actions américaines plutôt que des actions japonaises ou européennes).

> Choisir la répartition sectorielle au sein des classes d'actifs, des instruments financiers et des émetteurs susceptibles de générer de la performance (exemple : santé, aéronautique).

Évidemment tout l'enjeu de la gestion de portefeuille ou du trading est de déterminer à l'avance, la classe d'actifs ou le secteur qui va le plus performer dans les jours, les mois ou les années à venir. Le Machine Learning peut ainsi devenir un véritable atout pour essayer de prévoir l'évolution des marchés financiers.

2.1.2 Application du Machine Learning à la gestion de portefeuille

En gestion d'actifs, le Machine Learning s'avère d'une grande utilité pour prédire le cours des actifs. Il faut ainsi voir les marchés financiers comme une immense toile d'araignée. Tous les actifs sont liés entre eux et le moindre mouvement d'un actif a des répercussions sur un certain nombre d'autres actifs. C'est ce qu'on appelle la corrélation. Les actifs peuvent ainsi être corrélés positivement ou négativement. Les actifs d'un même secteur et/ou d'une même région sont souvent corrélés positivement.

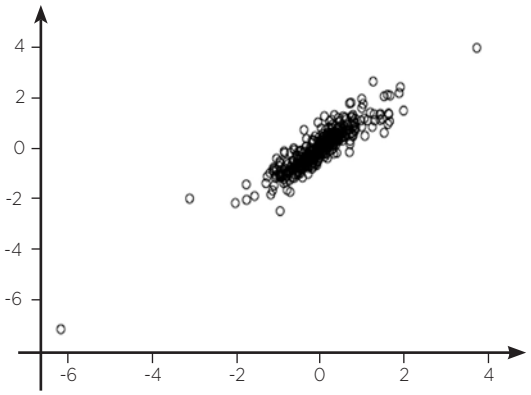
« DIFFÉRENCE ENTRE TRADING ET GESTION DE PORTEFEUILLE »

Pour comprendre l'intérêt du Machine Learning dans le domaine de la gestion d'actifs, il est important de faire la différence entre le trading et la gestion de portefeuille, l'utilisation du Machine Learning pouvant diverger selon les 2 cas. Le trading consiste en l'action d'investir afin de faire des profits à très court terme. Les décisions doivent être prises quasiment instantanément et les ordres passés immédiatement, c'est pour cela que les traders disposent d'un algorithme leur permettant de procéder à des achats et ventes automatiquement.

De ce fait, toutes les décisions sont prises par le biais dudit algorithme, le Machine Learning va donc constituer un élément essentiel du trading (mais qui a aussi ses limites).

La gestion de portefeuille, contrairement au trading, consiste en l'action d'investir avec l'objectif de générer des bénéfices à plus ou moins long terme. Le Machine Learning va, dans ce cas, plutôt avoir un rôle de robot-advisor[4], c'est-à-dire de conseiller les gérants sur les décisions à prendre grâce à des prédictions sur le marché financier (et l'analyse financière des entreprises et émetteurs). Cependant c'est le gérant qui a le dernier mot en ce qui concerne son portefeuille.

Figure 5. Corrélation linéaire entre les actions Société Générale et BNP Paribas



Prenons l'exemple de deux banques françaises, la Société Générale et BNP Paribas. Les données utilisées pour observer la corrélation sont le prix des actions du 01/01/2016 au 01/06/2017.

Les points ont tendance à former une ligne, permettant donc de déduire que ces deux actifs sont corrélés positivement. Cela se confirme, par ailleurs, par un coefficient de corrélation égal à 0,911.

De nombreuses données possibles

Il existe de nombreuses données possibles pour essayer de déterminer le cours futur d'un actif. Ces données peuvent être numériques, comme des séries de prix d'actions, de taux et d'autres indicateurs macroéconomiques. Elles peuvent également être issues de sources plus diverses : articles de presse, états financiers, publications sur les réseaux sociaux et annonces officielles grâce à l'analyse de texte, mais aussi, de manière plus large, au travers des résultats de navigation sur internet ou d'images.

En effet les prix des actions sont, à titre d'exemple, très sensibles à l'image de l'entreprise d'où l'importance des réseaux sociaux.

Utilisation d'un arbre de décisions

L'arbre de décisions est l'une des méthodes les plus utilisées pour la prévision de rendement d'actifs financiers.

Ci-dessous l'exemple d'Airbus. La question est de savoir s'il est pertinent d'investir dans des

4. Plateformes délivrant des conseils financiers ou permettant la gestion automatisée d'un portefeuille d'actifs.

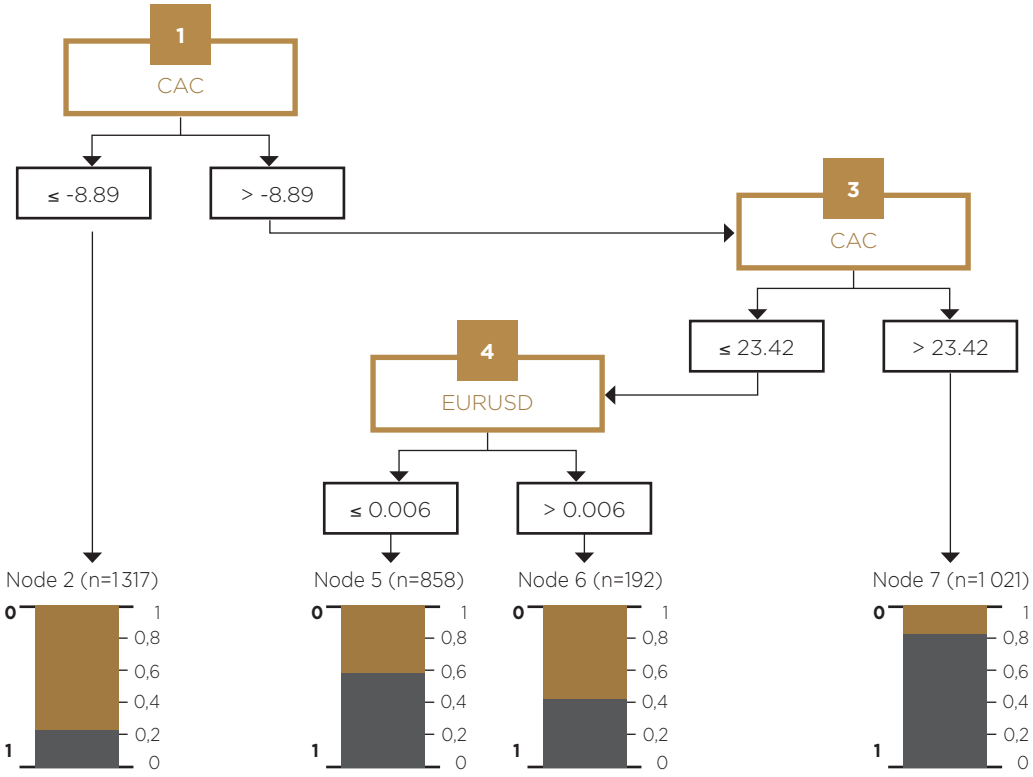
actions Airbus, compte tenu des informations suivantes :

> Des indices boursiers : le CAC40, le Dow Jones et l'Euro Stoxx 50 ;
> Des compagnies aéronautiques : Boeing, Bombardier ;
> Une compagnie aérienne : Air France.

Le jeu de données sera constitué de données journalières, comprises entre le 01/12/2003 et le 31/05/2019.

Ces données sont séparées volontairement en deux parties, l'une pour déterminer le modèle (par exemple 90% des données) et l'autre pour l'évaluer (les 10% restant). En utilisant l'algorithme C5.0 sur R[5], nous obtenons l'arbre de décisions suivant :

Figure 6. Arbre de décisions – Rendement Airbus



Le chiffre « 1 » correspond à un rendement positif et donc à un achat tandis que le « 0 » a un rendement négatif.

Le modèle se base uniquement sur les rendements du CAC40 et du taux de change EUR/USD pour prédire le rendement de l'action d'Airbus.

Ainsi si le rendement du CAC40 est inférieur à -8,89, le rendement de l'action d'Airbus a de fortes chances d'être négatif. Au contraire s'il est supérieur à 23,42, l'action d'Airbus aura sans doute un rendement positif.

Entre les deux, il faut utiliser le rendement du taux de change EUR/USD pour prédire celui de l'action d'Airbus.

En utilisant les données de test, il est alors possible de générer le tableau de validation croisée suivant :

5. Langage de programmation et logiciel libre destiné aux statistiques et à la science des données.

Tableau 1. Tableau de validation croisée – Rendement Airbus

	Prédiction		
Réalité	**0 (rendement négatif)**	**1 (rendement positif)**	**Total en ligne**
0 (rendement négatif)	131 (soit 34,7% de l'échantillon)	35 (soit 9,3% de l'échantillon)	166
1 (rendement positif)	72 (soit 19,1% de l'échantillon)	139 (soit 36,9% de l'échantillon)	211
Total en colonne	203	174	377

Le modèle est assez fiable lorsqu'il prédit que le rendement de l'action Airbus sera positif, évitant ainsi tout investissement à perte. En revanche, il existe un risque de passer à côté de bonnes opportunités : en effet dans 19% des cas, le modèle prédit une chute du cours de l'action, alors que finalement celui-ci monte.

Il est évidemment nécessaire d'être en possession de beaucoup plus de données, que ce soit en nombre de facteurs explicatifs ou de volume d'observations, pour avoir un modèle performant.

Utiliser les arbres de décisions devient une véritable aide pour les gérants d'actifs. Tout en gardant leurs convictions personnelles, ils peuvent savoir ce qui est statistiquement le plus probable.

Cependant, les rendements étant corrélés aux risques, il s'agit alors d'être capable de les mesurer pour mieux s'en protéger.

2.2 LE RISQUE DE CRÉDIT : NOTATION ET SCORING

La constitution d'un portefeuille de marché est un exercice complexe. Le recours au Machine Learning représente un atout d'envergure afin de pouvoir identifier au mieux le risque de crédit associé. Si l'utilité de ce procédé a été démontrée ci-avant, il s'avère que les éléments analysés reposent sur des informations fournies telles que les notations de titres corrélées à des situations de marché. Cependant, si les notations fournies par l'ensemble des grandes agences sont utilisées, il est également intéressant d'observer comment ces dernières sont déterminées, et surtout, comment elles vont être impactées par le Machine Learning.

2.2.1 La dette souveraine

Les dettes souveraines, matérialisées par des obligations, représentent une catégorie de produits financiers prisés des investisseurs et autres gérants de fonds. Entre 2014 et 2019, l'encours obligataire de la dette mondiale a été multiplié par 2,5 pour atteindre les 250 000 milliards de dollars.

La notation des obligations par les principales agences de notations (telles que Moody's, Fitch et S&P) devient donc primordiale pour les investisseurs en quête d'informations qualitatives. Le processus de notation est cependant une analyse longue et complexe. Le recours à l'intelligence artificielle et aux algorithmes de Machine Learning vient ici simplifier mais également affiner le travail des analystes (sans pour autant le remplacer).

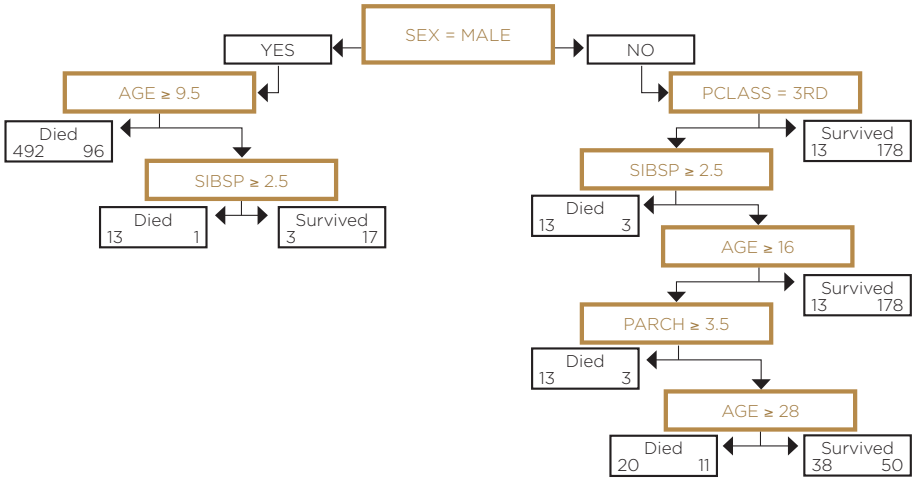
Ainsi, afin de prédire les notations des dettes souveraines (exemple : AA pour la dette française), plusieurs critères vont être observés pour décrire et analyser la capacité des emprunteurs souverains à tenir leurs engagements.

LE MODÈLE DES ARBRES DE DÉCISIONS

Le modèle des Arbres de décisions se présente comme une structure de données hiérarchisée avec, à chaque nœud, une décision devant être prise pour prévoir le résultat d'une classe.

Le but va généralement être de répartir une population d'individus possédant les mêmes caractéristiques selon différentes variables descriptives (exemple : âge, sexe, taille, etc.). Le choix des critères, des caractéristiques, des variables permettant la partition au fur et à mesure des données est fondé sur des indicateurs statistiques. C'est-à-dire qu'itération après itération, les choix sont de plus en plus restreints pour finalement n'avoir qu'une solution possible, qui est la décision finale ou le résultat final.

Ci-dessous l'exemple d'un arbre de décisions montrant qu'à chaque nœud se créent de nouvelles classes homogènes. Ici, nous cherchons la probabilité de vie ou de décès selon différents critères (élément fortement observé dans le domaine de l'assurance).



Les arbres de décisions sont des outils populaires dans le domaine du Machine Learning car ils fournissent des règles de décision claires et donc faciles à interpréter. Ils ont l'avantage d'être simples dans leur utilisation car ils évitent de procéder à de grandes transformations de données en entrée pour être utilisés. De même, il n'existe pas de paramètres à estimer contrairement à d'autres modèles. Ces éléments, combinés à un faible coût en ressources de calcul, présentent une véritable plus-value à l'utilisation.

Les arbres de décisions permettent, par ailleurs, de partir d'un résultat final et de trouver les règles qui ont permis d'atteindre ce résultat. Ceci se fait en prenant pour point de départ un jeu de données auquel on veut appliquer l'algorithme fourni par l'arbre de décisions. Cependant, quelques inconvénients existent à ce modèle. En effet, en fonction des volumétries de données observées, nous ne ferons pas toujours face à une classification optimale avec des approximations existantes.

Autre problématique rencontrée et quelque peu surprenante, il existe un phénomène de surapprentissage venant biaiser les résultats générés. Autrement dit, le modèle devient tellement efficace et spécifique à l'échantillon d'entrainement qu'il ne s'adapte pas à de nouvelles données et perd donc en précision et robustesse.

Voici à titre d'exemple quelques indicateurs analysés :

> **Indicateur 1 :** Dette externe / Exports
> **Indicateur 2 :** Balance externe
> **Indicateur 3 :** Taux d'inflation
> **Indicateur 4 :** Equilibre fiscal

La liste globale des indicateurs observés va constituer le jeu de données à analyser. L'avantage de ces derniers est qu'ils sont reconnus comme fiables et reposent généralement sur des mesures économiques solides.

Dans le cadre des notations souveraines, le modèle aujourd'hui privilégié est le modèle ACRP (Automated Credit Rating Prediction), lui-même basé sur le modèle du réseau de neurones[6] artificiels. Concrètement, il va analyser les différentes relations entre les variables explicatives (les indicateurs mentionnés ci-avant) et la variable à expliquer (la note de la dette) pour :

> Classifier (soit analyser un grand nombre d'individus, ici nos dettes souveraines, que l'on cherche à répartir en catégories, nos notations) ;
> Effectuer une régression (soit une modélisation établissant une ou des estimations futures à partir de données issues du passé). Au travers du modèle généré, il est alors possible de déterminer une note qui sera alors attribuée à une dette souveraine.

Une étude coréalisée par le Luxembourg Institute of Science and Technology et l'Université de Saarland (Allemagne) a ainsi pu fournir les pourcentages de précision dudit modèle pour ces deux formes :

Tableau 2. Comparaison de deux formes du modèle ACRP

Critères de performance	Réseaux neuronaux (ACRP) Classification	Réseaux neuronaux (ACRP) Régression
Notes souveraines correctement classifiées (en %)	40,4	34,6
Notes souveraines correctement classifiées (en %) avec un degré d'écart toléré	63,6	68,9
Notes souveraines correctement classifiées (en %) avec deux degrés d'écart tolérés	80,4	87,3
Notes souveraines correctement classifiées (en %) avec trois degrés d'écart tolérés	87,6	96,7

La qualité des résultats est variable. En effet, le modèle, tant en classification qu'en régression, ne donnera la note précise que dans 40,4% et 34,6% des cas, ce qui est bon mais pas suffisant.

Cependant, dès lors que l'on accorde une certaine marge d'erreur avec un ou plusieurs degrés tolérés dans la notation (un degré étant l'écart séparant une notation AAA d'une notation AA+), les résultats s'améliorent sensiblement (près de 9 cas sur 10 bien classés à 2 échelons près).

Le modèle de réseaux neuronaux se révèle d'une très grande efficience avec une note proche de la note réelle, en faisant un excellent indicateur pour les analystes qui viendront, dans certains cas, affiner le résultat.

Cette forte efficacité fait du modèle de réseaux neuronaux un modèle de référence pour cette typologie d'émetteurs de dette, cependant cela s'applique-t-il à d'autres typologies ?

6. Unité effectuant un traitement sur les données qu'il reçoit. Charge aux neurones d'envoyer ou non le résultat de son traitement aux neurones suivants.

LE MODÈLE DES RÉSEAUX NEURONAUX

Les réseaux neuronaux sont un modèle de Machine Learning qui s'inspire du fonctionnement de notre cerveau en créant des neurones, c'est-à-dire de petites unités recevant et traitant des informations. Chaque neurone effectue son propre traitement, qui est l'application d'une fonction dite d'activation, sur une somme de données qu'il reçoit en entrée et envoie ensuite ces données traitées au neurone suivant. Ainsi de suite jusqu'au neurone final qui donne le résultat du modèle.

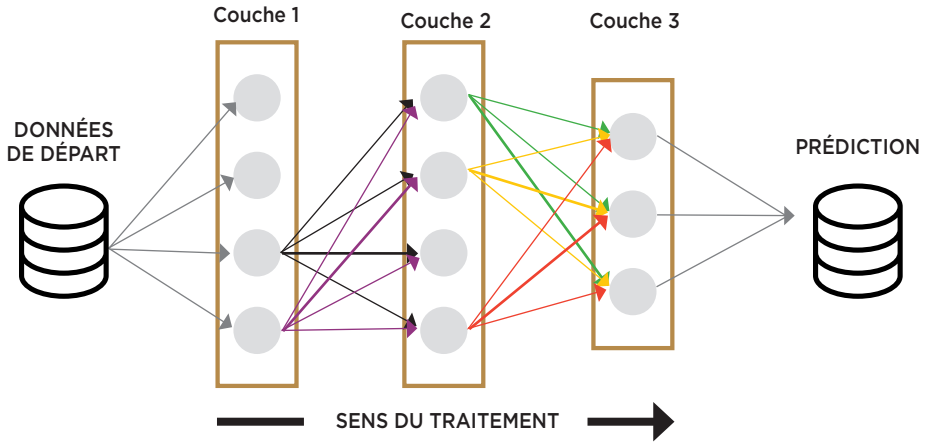
STRUCTURE DU MODÈLE

Les modèles usuels de Machine Learning ont recours à une multitude de neurones. Afin d'avoir un modèle opérationnel, l'organisation des neurones est un impératif pour obtenir une chaîne globale de traitement qui fournira une prédiction à partir de données injectées en entrée. Cette organisation est appelée « architecture[7] » du modèle. Une architecture est composée de plusieurs couches successives, qui sont des regroupements de neurones. L'ordonnancement des couches est important dans un modèle car le traitement des données se fait en fonction de cet ordre. Les données d'entrée passent par la première couche, puis par la seconde, et ainsi de suite. Ainsi un jeu de neurones de départ peut donner naissance à des réseaux différents selon l'architecture retenue car la constitution des couches et leur ordonnancement vont produire des traitements et donc des résultats différents. Trouver la bonne architecture est un enjeu majeur de l'implémentation d'un réseau de neurones en Machine Learning. Il n'existe pas, pour l'instant, de technique réellement définie pour trouver la structure idéale en fonction du problème traité.

CONNEXIONS NEURONALES

La passation d'informations d'une couche à une autre se fait via les « connexions neuronales ». Chaque neurone d'une couche va envoyer des informations aux neurones de la couche suivante. De la même façon que les connexions entre neurones dans le cerveau sont plus ou moins fortes selon l'utilisation et la pertinence de la connexion pour effectuer une tâche spécifique, les connexions en Machine Learning vont avoir des valeurs différentes dans la production du résultat du modèle.

Fonctionnement général d'un réseau de neurones



7. *Structure du modèle, définit la façon dont les neurones sont organisés afin de traiter les informations.*

LE MODÈLE DES RÉSEAUX NEURONAUX (SUITE)

Mathématiquement, les valeurs des connexions d'un neurone à un autre et d'une couche à une autre sont représentées par des poids. Ce sont ces poids qui sont utilisés pour pondérer la somme des informations qu'un neurone reçoit en entrée. Ces poids vont augmenter ou diminuer les informations reçues par les neurones et donc influencer le résultat du modèle.

Pour résumer, un réseau neuronal est un modèle très flexible de Machine Learning car il est possible d'arranger les neurones, c'est-à-dire les unités de traitement du modèle, comme nous le souhaitons. Une analogie peut être faite entre un réseau de neurones et une chaîne de production industrielle. Chacun a sa tâche à accomplir pour fabriquer un produit fini qui est ici une prédiction.

2.2.2 La dette corporate

L'analyse d'une dette d'Etat diffère totalement de celle d'une dette corporate. En effet, si les deux se basent sur des ratios, ces derniers sont bien différents, de même que les indicateurs externes pouvant présenter un impact sur leurs résultats.

Afin de pouvoir donner une indication fiable et indépendante, les entreprises du monde entier vont solliciter et payer des agences pour obtenir une notation de leur capacité à honorer le règlement de leurs dettes. Précieuse pour les investisseurs, cette note est générée suite à une longue et très complète analyse de ladite entreprise.

Afin d'être en mesure de proposer leurs coûteux services (de l'ordre du demi-million d'euros pour noter une entreprise) au plus grand nombre et pour obtenir des informations complémentaires au travail des analystes, ces agences vont avoir également recours aux modèles de Machine Learning (tout comme pour les dettes souveraines).

Combinés aux informations historiques des agences, de nombreux ratios vont servir d'éléments de référence. En voici quelques exemples :

> Dettes de long-terme / capital total investi ;
> Ratio de dettes ;
> Résultat opérationnel / capitaux reçus ;
> Marge brute.

Là aussi, le modèle ACRP, basé sur le modèle du réseau de neurones artificiels est utilisé et privilégié par les agences de notation. Cela s'explique en particulier par sa redoutable efficacité.

Afin d'illustrer le propos, quatre jeux d'entreprises pour lesquelles nous détenons les informations financières suffisantes ont été utilisés :

- USA-A : comprenant 265 entreprises américaines et uniquement 5 ratios analysés ;
- USA-B : comprenant 265 entreprises américaines et uniquement 16 ratios analysés ;
- TAIWAN-A : comprenant 75 entreprises taiwanaises et uniquement 5 ratios analysés ;
- TAIWAN-B : comprenant 75 entreprises taiwanaises et uniquement 16 ratios analysés.

Après application du modèle ACRP, les résultats suivants sont obtenus :

Tableau 3. Résultats du modèle ACRP

Jeu de données	Réseaux neuronaux (ACRP) Classification
USA-A	78,87%
USA-B	80,00%
TAIWAN-A	79,73%
TAIWAN-B	77,03%

La classification des entreprises est convaincante, néanmoins il est surprenant de voir une précision plus importante pour TAIWAN-A que pour TAIWAN-B alors que ce dernier dispose d'un plus grand nombre de données à observer.

Plusieurs hypothèses peuvent expliquer cela :

> Les dépendances entre les variables ne sont pas prises en compte explicitement par le réseau et la corrélation entre variables n'est plus un indicateur satisfaisant de leur dépendance ;

> Les nouvelles données d'entrées n'ont pas véritablement de lien avec la cible, pouvant détériorer la performance de l'algorithme. De même, de nouvelles valeurs d'entrées qui diffèrent de façon significative de celles qui ont été utilisées pour l'apprentissage du réseau peuvent dégrader les résultats générés (il s'agit du phénomène d'extrapolation) ;

> Un terme a été ajouté à la fonction d'erreur de l'algorithme pour pénaliser des valeurs de poids jugées trop importantes. Une modération des poids mal calibrée peut nuire à la performance du réseau en encourageant le sous-ajustement, ceci détériorant la précision des résultats attendus ;

Tableau 4. Résultats du modèle ACRP avec une faible marge d'erreur tolérée

Jeu de données	Réseaux neuronaux (ACRP) avec un degré d'écart toléré
USA-A	97,74%
USA-B	98,44%
TAIWAN-A	91,89%
TAIWAN-B	92,24%

Si on laisse une faible marge d'erreur tolérée (ici, un degré de notation), nous obtenons les résultats plus intéressants :

> Les résultats sont bien plus probants que pour les dettes souveraines, avec dans le cas du jeu de données USA-B, une efficacité des plus importantes venant illustrer la pertinence du modèle ;

> Encore une fois, la notation fournie ne sera pas suffisante, cependant, elle sera un indicateur très fort pour l'analyste qui n'aura que peu d'impact dans la décision finale (dans plus de 9 cas sur 10, la note est bonne ou affiche un seul degré d'écart avec la note définitive) ;

> Si, dans le cas des dettes corporate, le modèle est plus pertinent, cela tient notamment du fait que la notation d'une dette souveraine prend en compte de nombreux critères difficilement quantifiables tels que la politique intérieure, extérieure, les conflits commerciaux ou autres guerres.

2.3 LE RISQUE DE MARCHÉ

Le risque de marché représente un enjeu majeur du monde financier. Nous le caractérisons souvent par le risque de pertes résultant de l'évolution des prix du marché et des valeurs des actifs observés. Ce risque est particulièrement suivi d'un point de vue règlementaire. Il s'agit de l'un des volets mis en avant par le comité Bâlois (parmi d'autres) ainsi que l'un des principaux motifs de l'existence d'EMIR - European Market Infrastructure Regulation (se concentrant sur les produits dérivés).

En effet, les produits de marché représentent des montant colossaux et peuvent être à l'origine d'impacts considérables pour les gestionnaires d'actifs (les encours gérés pesaient 4 000 milliards d'euros en France en 2018).

L'objectif de la Direction des Risques d'un établissement financier sera alors de réduire autant que possible le risque de perte en capital sur les investissements effectués, tant pour compte propre que dans le cadre d'opérations visant à alimenter SICAV et autres OPCVM qui seront ensuite vendus à la clientèle. Focalisons-nous ici sur le risque action.

2.3.1 Diminution du risque avec le Machine Learning

Comme expliqué ci-avant, les algorithmes de Machine Learning ont la particularité d'être friands de grandes quantités de données, qui plus est de données de qualité.

Effectivement, les jeux de données issus des marchés présentent l'avantage d'être facilement accessibles tout en étant massifs et de bonne qualité car produits par des processus industrialisés peu sensibles aux erreurs de manipulation ou aux saisies.

Les algorithmes seront alors en mesure de procéder à l'étude approfondie de ces ensembles afin de distinguer les titres entre eux, créer des sous-ensembles et les classifier. Certains algorithmes, dits algorithmes de prédiction tenteront même de prédire de nouvelles données.

En ressortiront les titres présentant les risques les plus importants (variation de la volatilité, à court terme comme à long terme) qui pourront ainsi être écartés des placements réalisés ou même de ceux prévus (en fonction des politiques d'investissement).

Il s'agit d'un véritable outil venant compléter les analyses réalisées par les Risk Managers qui pourront les intégrer dans leurs estimations et alors accéder à des informations jusqu'alors inaccessibles. Car là où l'humain se perd face à une quantité massive de données, l'algorithme s'épanouit, apprend et s'améliore.

Il est ainsi de raison de s'interroger quant à l'efficience des modèles existants. Sont-ils tous pertinents face à cette typologie d'utilisation ? Lequel présente les meilleurs résultats ?

A ce titre, prenons 3 modèles populaires :

> Le Naïve Bayes,
> Les arbres de décisions,
> Les réseaux neuronaux.

Ces trois modèles affichent des caractéristiques spécifiques dans leur fonctionnement, mais également dans leur simplicité d'utilisation, et donc dans leurs résultats.

Pour illustrer cette comparaison, nous nous reposerons sur les résultats issus d'une étude réalisée par la Southern Illinois University présentant des éléments chiffrés venant illustrer nos propos.

2.3.2 Les données observées

Afin de pouvoir distinguer les modèles entre eux, deux jeux de données différents portant sur les marchés Européens et Japonais ont été étudiés. Ces deux jeux sont eux-mêmes composés de 59 caractéristiques financières différentes et portent respectivement sur 4 788 sociétés européennes et sur 3 644 sociétés japonaises (permettant de disposer d'une conséquente base à analyser).

Le choix de ces jeux de données a été effectué car ils représentent avant tout une large gamme de chiffres différents (ratio de dette, valeur de l'entreprise, cash disponible, etc.), le tout sur deux marchés distincts. L'un des principaux objectifs des algorithmes choisis sera de procéder à la reconnaissance cachée ou indirecte des relations entre les caractéristiques étudiées (et que l'analyste ou le data-scientist a du mal à trouver).

L'efficacité de cette reconnaissance est l'un des points d'attention fondamentaux de la comparaison des algorithmes analysés.

Si la volatilité des titres est observée, d'autres éléments telles que les caractéristiques financières des sociétés comme par exemple le taux de réinvestissement, l'EBIT ou l'EBITDA le seront également.

L'efficacité des 3 algorithmes dépend directement de leur facilité à distinguer et comprendre l'ensemble des relations entre les itérations d'entraînement sans connaissances préalables des dépendances économiques et financières des caractéristiques.

Autrement dit, nos 3 modèles vont, au fur et à mesure des différents entrainements, devoir trouver des liens entre les différentes caractéristiques observées sur les sociétés. Celui ou

ceux qui en trouveront le plus seront à même de maximiser leur efficience de classement.

Ainsi, les jeux de données choisis représentent un bon champ expérimental pour tester l'efficacité des algorithmes de Machine Learning utilisés et analyser le niveau de dépendance entre les propriétés financières choisies.

2.3.3 Les résultats d'analyse

Afin d'entrainer les algorithmes, ces derniers ont été testés en utilisant 10 expérimentations réalisées sur les 2 jeux de données. En ressortent les résultats suivants (cf. figure 7).

Le modèle algorithmique des arbres de décisions montre une meilleure efficience en termes de classification que ses comparses (bien que les résultats des Réseaux Neuronaux soient honorables pour la classification des sociétés japonaises).

Néanmoins, si l'entraînement est poussé pour être bien plus intensif avec une bien plus grande récurrence d'expérimentations réalisées sur les

Figure 7. Performance des algorithmes après 10 entraînements

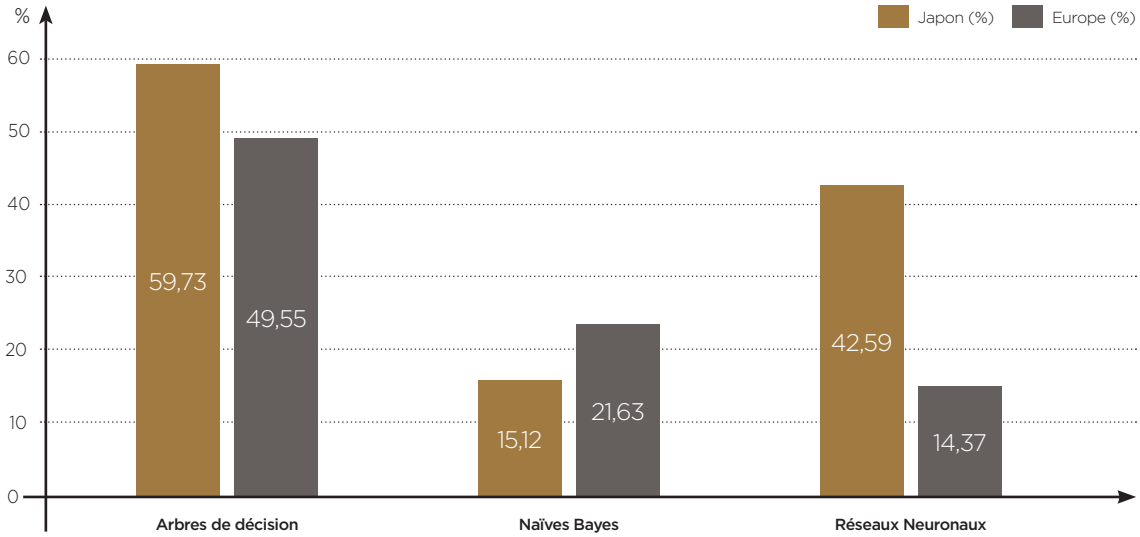
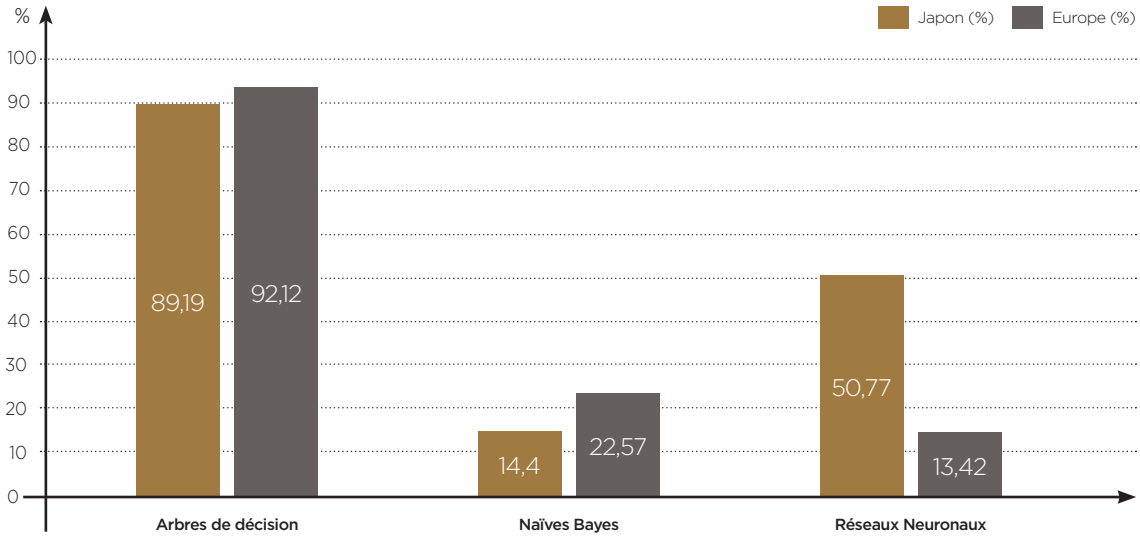


Figure 8. Performance des algorithmes après un entraînements intensif



2 jeux de données, les résultats obtenus sont alors différents (cf. figure 8).

Le résultat est sans appel, le modèle des arbres de décisions est bien plus précis dans sa classification, tant pour les sociétés japonaises qu'européennes.

Il est même constaté une régression pour certains modèles par rapport à un entraînement moindre.

Cela vient notamment prouver que tous les modèles ne sont pas égaux entre eux. Certains vont tirer leur épingle du jeu dans certaines situations tandis que d'autres non. L'exemple mentionné ci-dessus révèle la supériorité du modèle des arbres de décisions, modèle pourtant jugé moins complexe que celui des réseaux neuronaux.

Le Machine Learning peut ainsi être un outil des plus pertinents pour le risque de marché car il repose sur une variété et une quantité de données fournissant aux algorithmes la capacité à créer des liens.

La pertinence d'un algorithme se dégradera ou s'améliorera selon le contexte pour lequel il est utilisé. Au travers des résultats obtenus par les data scientistes, le Machine Learning viendra en aide aux analystes pour devenir complétement pertinent. Pour pouvoir pleinement révéler son potentiel au sein des différentes fonctions des banques et assurances, plusieurs défis restent à relever.

LE MODÈLE NAÏVE BAYES CLASSIFIER

Le Naïve Bayes Classifier est l'un des algorithmes de classification les plus simples. S'il est régulièrement utilisé pour la classification de textes, celui-ci possède un spectre d'utilisation large et peut être adapté à de nombreux domaines d'utilisation (notamment en finance).

Le théorème de Bayes repose sur un modèle probabilistique simple, à savoir de déterminer **« la probabilité de A sachant B ».**

$$P(A|B)=\frac{P(B|A).P(A)}{P(B)}$$

Exemple n°1 : Quelle est la probabilité que Julien ait une barbe sachant qu'il porte des lunettes ?

En l'état, ces deux critères ne sont pas nécessairement interdépendants[8]. Quel lien entre des lunettes et une barbe si ce n'est peut-être un intérêt esthétique ?

Néanmoins, avec l'exemple suivant, la relation change totalement.

Quelle est la probabilité que Julien porte des lentilles sachant qu'il a une mauvaise vue ?

L'interdépendance[8] entre les deux critères (la mauvaise vue et les lentilles) semble être plus forte que pour le premier cas (barbe et lunettes). Il devrait y avoir, logiquement, une probabilité plus grande dans le second cas.

C'est ce que va rechercher le modèle de classification Baïsien, mais avec une vision plus large. Car, en effet, s'il n'était observé jusqu'ici qu'une seule variable, l'algorithme de Machine Learning « Naïve Bayes Classifier » va prendre en compte plusieurs variables distinctes afin de les étudier indépendamment les unes des autres.

8. État de choses ou de personnes qui dépendent les unes des autres.

LE MODÈLE NAÏVE BAYES CLASSIFIER (SUITE)

Exemple n°2 : Le tableau ci-dessous indique le nombre de clients ayant un contrat de crédit à la consommation (amortissable ou renouvelable) ou un compte-courant auprès d'une banque et leur situations (Sain ou défaut).

Situation	Crédit à la consommation	Compte-courant	Totaux
Défaut	3250	6230	9480
Sain	20450	36540	56990
Totaux	23700	42770	66470

Le but dans ce cas est de calculer la probabilité qu'un client soit en défaut sachant qu'il a souscrit un crédit à la consommation. Le programme calcule la probabilité d'être en défaut et la probabilité d'être en défaut avec un crédit à la consommation pour obtenir le résultat final.

> La probabilité d'être un client en situation de défaut : **14,3%**

> La probabilité d'être un client en défaut détenant un crédit à la consommation : **4,89%**

Il en ressort :

> La probabilité que le client soit en défaut sachant qu'il possède un crédit à la consommation : **34,28%**

S'il est vrai que toutes les variables peuvent avoir une certaine dépendance les unes par rapport aux autres, ce modèle choisit d'ignorer cela, ce qui en fait à la fois sa force (simplicité et vitesse de traitement) et sa faiblesse (car il passe volontairement à côté de certaines informations).

Le Naïve Bayes Classifier est donc un modèle accessible et simple à mettre en place. Il présente surtout une rapidité d'exécution pouvant être appréciée dans le cadre de certaines études et ne nécessite pas une grande puissance de calcul. De même, celui-ci est prisé car accessible avec de « petits » jeux de données.

Si les résultats générés ne sont pas toujours les plus précis, ce modèle d'apprentissage supervisé se voit être plutôt pertinent avec de bons résultats de classification, parfois meilleurs que d'autres modèles plus complexes, et ce en dépit de son hypothèse de simplification.


control
option
command


Code File Edit Selection View Go Debug Tasks Window Help
EXPLORER
OPEN EDITORS
package.json edgeweb-workspace/reseller
typings.json edgeweb-workspace/reseller
tsconfig.json edgeweb-workspace/reseller
api.ts edgeweb-workspace/reseller/src/app/s...
EDGESTACK
edge-deployment
edge-workspace
edgeweb-workspace
admin
org
reseller
node_modules
src
app
datatypes
directives
pages
services
api
api.ts
index.ts
app
auth
brokers
data
dates
jwt
login
orgs
readings
user
widgets
app.component.html
import { Injectable, Inject, Injector } from '@angular/cor
import { Router } from '@angular/router';
import { AuthHttp, JwtHelper } from '../jwt/jwt';
import { Observable } from 'rxjs/Rx';
import { AuthHelper } from '../auth/auth';
import { RequestOptions, Headers, Response } from '@angula
import { Config } from '../../config';
import { appInjector } from '../app/app-injector';
@Injectable()
export class ApiService {
jwt: string;
private static parseResponse(res: Response) {
var info = res.headers.get('X-Message-Info');
var error = res.headers.get('X-Message-Error');
var success = res.headers.get('X-Message-Success');
return {
message: {
info: info,
error: error,
success: success
},
json: function() {
let json;
try {
json (local var) Exception: any
} catch (Exception) {
json = res.json();
}
return json;
}(),
res: res
};
}
constructor(
public auth: AuthHelper,
public authHttp: AuthHttp,

3.

LES DÉFIS DU MACHINE LEARNING

Le Machine Learning représente au même titre que d'autres innovations technologiques (telles que la Blockchain ou l'informatique quantique) une avancée à même de venir améliorer l'efficience des entreprises. Cette technologie n'en reste pas moins imparfaite à cause de nombreux défauts entachant son utilisation. Il n'est pas rare d'obtenir des résultats surprenants sans pour autant en comprendre la raison. Mais quelles sont exactement ces limites ?

3.1 LES PROBLÉMATIQUES DES MODÈLES

3.1.1 Le choix du bon modèle

Les modèles de Machine Learning sont nombreux et variés, capables de répondre à nos besoins. Cependant ils ne sont pas tous adaptés, loin de là. Comme dans toute pratique, certains outils n'auront qu'une fonction dédiée et ne pourront servir à autre chose.

Ainsi, choisir un mauvais modèle conduira à de mauvais résultats et potentiellement à de mauvaises décisions. Il est nécessaire de sélectionner celui qui sera le plus adapté, soit un modèle :

> Capable de procéder à des prédictions au travers des données utilisées, mais également avec de nouvelles données pouvant apparaître avec le temps ;
> Permettant d'obtenir le(s) résultat(s) désiré(s) (ex : l'application d'un modèle de régression pour une problématique de classification ne sera pas nécessairement adaptée) ;
> S'accordant avec la typologie des données observées (ex : il existe une différence forte entre du texte et des images) ;
> Prenant en compte les contraintes de ressources informatiques (puissance du serveur et taille de la mémoire dédiée) ainsi que le temps de calcul qui est toléré par les utilisateurs (certains modèles vont générer des résultats plus rapidement que d'autres) ;
> Susceptible de fournir le niveau de précision attendu (nous avons pu comparer précédemment plusieurs modèles avec une précision variable).

Une véritable pré-étude va alors être nécessaire afin de pouvoir anticiper ces éléments et éviter un choix inadapté. Il conviendra pour cela de déterminer plusieurs caractéristiques :

Figure 13. Que faut-il analyser pour choisir le bon modèle ?

Le contexte et les motivations relatives au besoin	**L'objectif lié à l'utilisation du Machine Learning**	**Le jeu de données mis à disposition et sa forme**	**Les limitations (techniques, temporelles, savoir-faire)**

Il s'agit en effet d'un minimum requis pour éviter un choix inadéquat. Au travers de la définition de ce cadre, une meilleure visibilité sera fournie aux équipes de Data-Scientistes leur autorisant un travail d'une plus grande efficacité et surtout en adéquation avec les attentes métiers à l'origine de la requête. S'il s'avère que le choix du modèle représente un véritable impératif, il n'en demeure pas moins que d'autres problématiques, peut-être plus surprenantes, existent.

3.1.2 Le cas du surapprentissage

L'usage de modèles algorithmiques de Machine Learning n'est pas aussi aisé qu'il est coutume de le penser. Problématiques d'alimentations en données, qualité de ces dernières et choix du bon modèle sont tout autant de contraintes à considérer pour utiliser cette technologie de manière optimale.

Il reste néanmoins un cas particulier concernant les algorithmes eux-mêmes : le « surapprentissage » (en anglais « overfitting »).

Il s'agit d'un cas particulier dans lequel le modèle est tellement efficace qu'il ne s'adapte pas à de nouvelles données lui étant inconnues et perd sa précision.

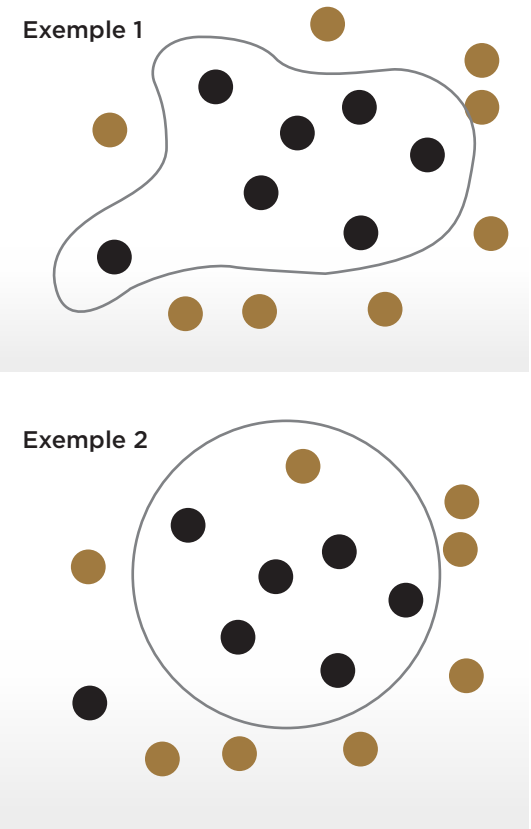
Il se perd lui-même dans son analyse devenue trop complexe et finit par ne plus distinguer les corrélations utiles de celles qui ne le sont pas (que l'on peut appeler « bruit »). Il est devenu finalement trop complexe par rapport à la réalité qu'il essaie de représenter.

Cette situation apparaît dans certains cas de figure dès lors que l'on entraîne trop notre algorithme.

Dans l'exemple 1, on constate que le modèle a bien ciblé les points noirs tout en veillant à ne pas intégrer les points dorés dans ses résultats. Par contre, l'exemple 2 est moins précis. Du fait d'un « surapprentissage », le modèle présente une qualité moindre dans ses résultats avec un point noir ignoré et surtout un point doré intégré.

Un bon modèle est un modèle qui est à la fois adapté au jeu de données et aux objectifs fixés mais également équilibré au travers d'un entrainement adéquat. Il sera alors possible

Figure 14. Exemples de cas de surapprentissage



d'éviter qu'il soit submergé de données et n'arrive plus à les interpréter correctement.

La sensibilité des algorithmes de Machine Learning constitue un véritable challenge pour les maitriser au mieux, d'autant plus que leur obscurité n'a pas encore été totalement décryptée.

3.1.3 L'obscurité des réseaux neuronaux

Le modèle de réseaux neuronaux attire par ses nombreux avantages. Les résultats obtenus sont de grande qualité avec de meilleures performances que d'autres algorithmes (comme les arbres de décisions).

Ses possibilités d'utilisation sont multiples et il permet de travailler avec des données incomplètes ou bruitées mais n'est malheureusement pas parfait.

Si les résultats obtenus sont certes pertinents, ils en deviennent difficilement explicables, même pour leurs concepteurs. Les réseaux de neurones fonctionnent comme une « boîte noire » contrairement à d'autres modèles comme les arbres de décisions qui ont une logique de raisonnement claire.

Si pendant de nombreuses années, l'opacité relative à ces modèles ne posait pas de réels problèmes (il s'agissait de sujets de laboratoires), leur utilisation dans le monde de l'entreprise, et plus particulièrement de la finance, interpelle.

En effet, les domaines bancaires, assurantiels ou encore celui de la gestion d'actifs sont tous soumis à des règlementations fortes de la part d'autorités externes. Au travers de l'utilisation de modèles dits « boites noires », un établissement pourra se retrouver en situation d'inconfort lorsqu'il lui sera demandé d'en expliquer le fonctionnement et de justifier certaines prises de décision.

La difficulté à auditer les réseaux neuronaux lors de contrôles de la part des autorités régulatrices (exemple : l'ACPR – Autorité de Contrôle Prudentiel et de Résolution) se transforme en un risque pour une banque.

Comment justifier que les résultats utilisés, voire même communiqués au régulateur, sont exacts et adaptés ?

Si l'utilisation de ces modèles en est à ses débuts, il est préférable de garder une certaine mesure quant à l'application de ces derniers. De nombreux scientifiques tentent aujourd'hui de travailler sur ce sujet, à commencer par Laurence Devillers, chercheuse au CNRS et auteur de l'essai « Des robots et des hommes », « *s'il est important de rendre ces systèmes plus robustes, il faut aussi expliquer comment ils fonctionnent, et garder à l'esprit qu'ils apprennent sans comprendre, donc qu'ils décident sans comprendre.* ».

3.2 LE CHALLENGE DE LA COLLECTE DES DONNÉES

L'existence de nombreuses typologies de modèles ouvre le champ des possibles. Si les utilisations sont multiples et adaptables aux besoins, le monde financier n'a su s'intéresser à ces capacités que tardivement, ou de manière réduite. Il n'en reste pas moins que le Machine Learning représente une véritable opportunité pour les établissements bancaires d'optimiser leur acquisition de clients et leurs actions commerciales.

Les différents modèles présentés vont être utilisés pour répondre aux besoins. L'alimentation en données de qualité deviendra donc l'enjeu principal des banques qui devront réussir à obtenir lesdites données, en quantité, pour être en mesure de libérer le plein potentiel de ces puissants algorithmes.

Il convient donc de s'interroger sur la manière dont les banques et autres établissements financiers vont être en mesure d'obtenir ces données, et surtout de voir comment celles-ci seront exploitées.

3.2.1 Le KYC : d'une contrainte nait une force

Le KYC, Know Your Customer, est une obligation réglementaire imposant la vérification de l'identité des clients d'un établissement.

Servant de base d'informations sur ces derniers, les données collectées sont utilisées dans le cadre de son activité de Conformité (ou Compliance) afin de respecter les directives européennes sur la lutte contre le blanchiment d'argent, le financement du terrorisme ou encore contre la fraude (LCB-FT).

Ce processus intervient à l'entrée en relation, mais également au cours de la vie de cette relation. L'objectif est de s'assurer de l'identité du client au travers d'informations vérifiables et certaines.

Figure 9. Processus d'identification KYC

Figure 10. Exemple de données recueillies :

Plusieurs problématiques ressortent de cette obligation :

- Elle est chronophage pour les établissements ;
- Elle est donc coûteuse ;
- Elle parasite l'expérience client (demande de nombreux documents justificatifs, allongement du temps d'entrée en relations, incompréhension client, etc.) ;
- Elle doit être suivie tout au long de la vie de la relation avec le client (mises à jour régulières) ;
- Il est nécessaire de pouvoir stocker ces informations, de manière ordonnée et sécurisée.

Très longtemps gardés sous forme de document papier, les documents collectés auprès du client ont par ailleurs nécessité un certain temps pour être numérisés par les établissements. En effet, le volume des éléments en possession représentait un réel challenge (tant en termes de logistique qu'en terme de stockage et de structurations des données).

Véritable facteur de coûts, le KYC est alors vu comme une contrainte ainsi qu'un frein aux relations commerciales et donc à la croissance du ***Produit Net Bancaire*** des banques.

Toutefois, le KYC permet d'obtenir une véritable base de données, riche, travaillée, qualitative et de taille non négligeable concernant la clientèle.

Une fois ces données retravaillées et structurées,

les possibilités seront multiples et permettront notamment aux équipes de Data Scientistes d'avoir recours aux modèles de Machine Learning afin de cibler certaines catégories de client pour plusieurs usages :

> Opérations commerciales dédiées (produits spécifiques) ;
> Détection de changements dans la vie du client (revenus, situation familiale) ;
> Personnalisation de la communication client (satisfaction) ;
> Diminution de la fraude.

Au travers de ces actions, l'établissement sera à même de contrebalancer les coûts engendrés par la constitution du KYC client par un PNB nouvellement généré ou par une réduction du coût du risque (sans oublier une satisfaction clientèle améliorée). Le client est connu, mais il est ainsi possible de le fidéliser davantage.

Le monde financier, et plus particulièrement le monde bancaire, ne s'est rendu compte que tardivement du véritable intérêt de la culture de la donnée client. Elle représente l'un des principaux défis de demain.

La collecte doit ainsi être simplifiée, complète, mais avant tout précise et industrialisée.

C'est notamment l'une des raisons pour lesquelles l'automatisation des processus de contrôle KYC doit être priorisé, tout comme la simplification de la collecte et du stockage des données, une voie sur laquelle la plupart des banques se sont engagées pour réaliser d'importants gains en efficacité opérationnelle. Cette automatisation peut être opérée en interne ou s'appuyer sur des partenariats externes.

C'est par ces procédés que l'établissement concerné sera alors en mesure de travailler à la création d'une véritable base de données clients, qualitative (car reposant sur des informations vérifiées et confirmées par des documents au caractère légal), mais également quantitative.

3.2.2 Le parcours client

Si nous avons pu constater que, par l'entrée en relation, nous disposons d'une importante quantité d'informations de qualité sur notre client, celles-ci ne représentent que la partie visible de l'iceberg.

Un établissement bancaire va, grâce à la simple typologie de produits qu'elle propose, recueillir des informations quantitatives et qualitatives sur le client et son profil.

L'ensemble des données (flux, crédits, débits, typologie d'opérations, mais aussi épargne) constitue des informations présentes et existantes dans les bases de l'établissement hôte des comptes clients. Très précieuses, ces informations permettent d'en apprendre beaucoup sur les habitudes du client (sous réserve qu'il soit consentant avec le stockage de ces dernières, en accord avec règles prônées par le RGPD[9]).

Que fait notre client de son argent ? A-t-il un profil économe ? Ou au contraire, est-il dépensier ? Où vont ses flux ? De nombreuses questions sont possibles et des réponses vont être ainsi centralisées à travers les flux et leur analyse. Cette centralisation va amener la banque à constituer d'importantes bases de données retraçant l'ensemble de sa connaissance du client, ce qui est devenu un enjeu majeur des dernières années.

Cette acquisition de données est d'autant plus pertinente qu'elle est fiable, liée à l'actualité récente du client et à ses objectifs ou projets. Au travers du comportement du client, il est alors possible de détecter ses centres d'intérêt, et de prédire ses besoins. Une multitude d'infor-

9. RGPD (ou Règlement Général sur la Protection des Données) : Règlement de l'Union Européenne ayant pour but de renforcer et d'unifier la protection des données pour les individus au sein de l'union Européenne.

CLASSIFICATION DES OPÉRATIONS BANCAIRES

Les opérations réalisées par la clientèle sont des données brutes qui nécessitent d'être retraitées afin d'être exploitables et interprétables.

Les clients des établissements bancaires sont aujourd'hui mis à contribution. En effet, l'ensemble des applications de consultations des comptes en ligne permettent désormais de classifier nos opérations.

Indiquer que votre ligne de relevé « Délices d'Asie » correspond à une activité « Restauration / Loisir » est désormais possible afin de mieux gérer votre budget.

En procédant à cette classification, vous allez aider un algorithme de Machine Learning à apprendre au fur et à mesure à quoi correspondent les opérations réalisées.

Progressivement, certaines opérations seront classées automatiquement. L'algorithme aura appris, à l'aide des utilisateurs de ce procédé, à quoi elles correspondent.

En soit, la classification facilite la compréhension des paterns de consommation des clients pour mieux prévoir et proposer les solutions adaptées.

Le détenteur du compte y trouve un avantage, à savoir de suivre la classification de ses dépenses pour mieux s'organiser dans sa vie de tous les jours. En contrepartie, l'établissement appréhende mieux ce que fait son client de ses fonds et en sait plus sur son profil et ses habitudes.

Les géants de l'industrie technologique tels que Google, Facebook ou Amazon avaient compris cela il y a plusieurs années. C'est notamment l'une des raisons pour lesquelles ils s'intéressent de près à la mise en place de systèmes de paiement ou à des alternatives bancaires.

La collecte de données représente l'essentiel de leur chiffre d'affaires. Obtenir ces nouvelles informations constitue un vecteur non négligeable de croissance aujourd'hui essentiellement détenu par le secteur financier.

mations vont donc être récoltées et observées pour venir alimenter les bases de données de l'établissement. En voici un exemple :

> E-Reputation et réseaux sociaux ;
> Données open-source (exemple: données démographiques) ;
> Epargne et mouvements bancaires centralisées par le biais d'agrégateurs de compte ;
> Informations externes anonymisées sur le comportement d'achat des consommateurs.

Chacune de ces données est, à titre individuel, une information observable intéressante cependant la combinaison de plusieurs de ces critères sera à l'origine de règles de classifications ou d'actions ciblées plus précises et adaptées.

Les algorithmes de Machine Learning adaptés au CRM, Customer Relationship Management, vont accroitre l'efficience des résultats, mais surtout, permettre d'obtenir de nouvelles typologies de résultats encore non observées avec la seule expertise humaine. Par exemple, il peut être détecté, au travers du score de Churn[10], la probabilité qu'un client transfère ses comptes au sein d'un établissement concurrent. Des actions pourront être ciblées pour maintenir la relation avec ledit client. L'objectif sera ainsi de

10. Score permettant d'analyser la fidélité d'une clientèle et l'impact d'actions sur celle-ci.

Figure 11. Conséquences d'une mauvaise qualité de donnée

procéder à l'optimisation des actions d'acquisition et de fidélisation client ou à la détection de comportements jugés « atypiques ».

Le maintien de parts de marché dans un environnement concurrentiel fort y trouve tout particulièrement son intérêt. Le changement d'établissement a en effet été facilité par la loi Hamon (pronant une mobilité bancaire simplifiée, automatisée et gratuite).

Avec un coût d'acquisition client en hausse, les banques auront un avantage certain à détecter les clients en partance vers la concurrence.

L'utilisation de modèles de Machine Learning y est ici intéressante. Moins statiques que les modèles classiques habituellement utilisés, ils permettent de s'adapter aux nouveaux comportements des consomateurs et évoluent dans le temps afin de les intégrer (exemple : un modèle de réseaux neuronaux en temps réel pourrait être à même de faire le lien entre l'utilisation de mots clés et de formulations utilisées par un client dans un chatbot pour générer une alerte sur un éventuel départ).

Identifier ces clients va permettre d'effectuer des opérations de reconquête de ces derniers, que cela soit au travers de nouvelles offres plus adaptées, des réductions tarifaires ou du redéveloppement de la relation.

3.2.3 L'enjeu de la qualité des données

Le Machine Learning s'apparente à un moteur auquel nous allons donner une fonction, capable de s'enclencher à partir de données. Il s'agit ici de son carburant. Un carburant d'autant plus précieux qu'il est difficilement obtenu (bien qu'il ait été vu de très nombreux moyens de s'en procurer) mais également de qualité variable : de nombreux établissements disposent de grandes quantités de données en leurs bases, cependant celles-ci ne seront pas toujours exploitables par les algorithmes.

Disposer d'une grande qualité de donnée va alors devenir un objectif prioritaire. Cela se traduit par un besoin simple, celui d'avoir une donnée exacte, claire, utile et fraiche (entre autres).

Des données de mauvaise qualité auront plusieurs impacts avec des conséquences parfois lourdes pour l'établissement qui les exploite.

Le principal risque reste néanmoins la génération de faux résultats tels que :

> De mauvaises segmentations ;
> Des estimations erronées ;
> Une tarification inadaptée ;
> Des engagements inappropriés.

Il en découle aisément des choix ou décisions pouvant aller à l'encontre de l'entreprise, générateurs de pertes ou d'insatisfaction client. C'est notamment pour ces raisons que le régulateur Français, l'ACPR (Autorité de Contrôle Prudentiel et de Résolution) procède à la surveillance de ces éléments au sein des divers établissements de la place financière. La campagne d'audit TRIM (Targeted Review of Internal Model) réalisée entre 2018 et 2019 en est un exemple significatif. Cette dernière avait pour objectif l'évaluation de l'adéquation des modèles internes mis en œuvre par les banques, mais aussi de la fiabilité de leurs résultats. Il en a résulté d'importantes recommandations propres à chaque établissement sur le déploiement des

modèles statistiques, une harmonisation des règles d'utilisation et de mise en œuvre des modèles à travers de nouvelles guidelines, et le déclenchement de vastes programmes de remédiation au sein des entités bancaires pour adapter leur dispositif de notation à toutes ces exigences.

Parmi celles-ci, la qualité des données est aujourd'hui perçue comme une menace suffisamment importante pour devenir l'un des piliers de certaines règlementations. C'est le cas de Solvabilité II, règlementation adaptée aux Assureurs ayant intégré cela comme l'un de ses trois principes fondamentaux. Des amendes du régulateur peuvent donc désormais porter sur cet élément.

Mais comment agir pour éviter ce risque ?

Il est tout d'abord primordial d'intégrer la qualité de données à sa politique interne de risques opérationnels. A l'heure où la donnée est utilisée en permanence, il s'agit de l'une des composantes importantes du patrimoine de l'entreprise et constitue un avantage concurrentiel.

Cela passe également par la mise en œuvre d'une gouvernance SI/Qualité des données avec l'application de plusieurs règles (notamment prônées par la circulaire BCBS 23911 du comité de Bâle) :

Figure 12. Règles prônées par la circulaire BCBS 23911

Fiabilité : garantir une image fidèle	**Lisibilité :** disposer d'une information homogène partagée	**Pérennité :** processus d'amélioration continue de la qualité	**Contrôle :** mise en place de processus de contrôle renforcés

La qualité de données devient donc un enjeu qu'il n'est plus possible d'ignorer pour le présent et l'avenir des entreprises. Des technologies de pointes telles que le Machine Learning ne peuvent être efficientes qu'à ces conditions. D'une faiblesse, il est désormais possible de faire une multitude d'opportunités.


Debug Tasks Window Help
Thu Aug 3 7:47 PM
api.ts — Edgestack
package.json
typings.json
tsconfig.json
api.ts
weakmap
No Results
import { Injectable, Inject, Injector } from '@angular/core';
import { Router } from '@angular/router';
import { AuthHttp, JwtHelper } from '../jwt/jwt';
import { Observable } from 'rxjs/Rx';
import { AuthHelper } from '../auth/auth';
import { RequestOptions, Headers, Response } from '@angular/http';
import { Config } from '../../config';
import { appInjector } from '../app/app-injector';
@Injectable()
export class ApiService {
jwt: string;
private static parseResponse(res: Response) {
var info = res.headers.get('X-Message-Info');
var error = res.headers.get('X-Message-Error');
var success = res.headers.get('X-Message-Success');
return {
message: {
info: info,
error: error,
success: success
},
json: function() {
let json;
try {
json (local var) Exception: any
} catch (Exception) {
json = res.json();
}
return json;
}(),
res: res
};
}
constructor(
public auth: AuthHelper,
public authHttp: AuthHttp,
Go Live
Ln 33, Col 17
Spaces: 4
UTF-8
LF
TypeScript
2.4.1

4.

CONCLUSION

Historiquement, le recours aux mathématiques et aux statistiques constitue un acquis de longue date au sein de l'industrie bancaire. En cela le Machine Learning constitue logiquement, au sein de la banque, une étape de plus dans le recours à la data science. Un usage d'autant plus d'actualité qu'il profite d'une forte incitation des pouvoirs publics pour accélérer la digitalisation de l'économie.

En l'état de son déploiement opérationnel, il apparaît quelque peu prématuré de qualifier le Machine Learning de révolution pour le monde de la finance. En réalité, il constitue une des facettes de la transformation numérique des services financiers, qui a vu la donnée devenir l'un de ses principaux leviers.

Les banques lèvent progressivement, les unes après les autres, tous les freins qui limitaient un recours élargi au Machine Learning. Ces freins touchaient à la donnée, devenue désormais très variée et disponible en grande quantité avec une qualité croissante (les autorités y veillent). Ils portaient également sur les puissances de calcul, qui bénéficient des derniers progrès technologiques en la matière. Ils concernaient enfin, l'outillage du data scientist, qui n'a jamais été aussi pléthorique, avec de nombreuses solutions logicielles, voire open source, disponibles pour tester les dernières méthodes de scoring. L'avènement du code Python a fini de consacrer une sémantique informatique quasiment internationale, donc homogénéisée et plus largement exploitable.

L'usage du Machine Learning reste malgré tout encore imparfait dans le monde de la finance. En effet, la majorité des modèles supervisés nécessite encore l'intervention humaine et n'a, pour les tâches les plus complexes, qu'un objectif simple, à savoir apporter une assistance dans l'analyse et la prise de décision. Il en sera de même pour l'emploi des modèles non supervisés, surprenants dans leur capacité de prédiction ou de classification mais pourtant toujours aussi obscures et difficiles à expliquer dans leur fonctionnement.

Et au-delà de son usage, des freins majeurs au déploiement du Machine Learning subsistent. Au premier rang desquels figurent la rareté (et donc le coût élevé) des compétences en matière de Machine Learning d'une part, et les obstacles culturels à déléguer certaines activités d'autre part. Ces obstacles ne sont pas rédhibitoires ; en réalité ils permettent de nuancer le miracle de l'intelligence artificielle dont la promesse ne présage pas la complexité de mise en œuvre. Et si finalement le Machine Learning ne constituait qu'une première étape pour constituer un terrain favorable à la révolution engendrée par un usage intensifié de l'IA ?

5.

RÉFÉRENCES DOMAINES D'EXCELLENCE CONTACTS

RÉFÉRENCES

Corporate Credit Rating using Multiclass Classification Models with order Information
Hyunchul Ahn and Kyoung-Jae Kim

Prediction of financial strength ratings using Machine Learning and conventional techniques"
Predicting Corporate Credit Ratings using Content Analysis of Annual Reports – A Naïve Bayesian Network Approach
Petr Hajek, Vladimir Olej, Ondrej Prochazka

A Comparison of Machine Learning Classifiers Applied to Financial Datasets
Pablo D. Robles-Granda and Ivan V. Belik

Data science : Fonametaux et études de cas
Eric Biernat, Michel Lutz

Banking Customer Segmentation and Machine
Sachin Jahagirdar
Learninghttps://www.linkedin.com/pulse/banking-customer-segmentation-machine-learning-sachin-jahagirdar/

How to Build a Product Recommendation System. Machine Learning Solutions
Michał Dyzma
https://www.netguru.com/blog/how-to-build-a-product-recommendation-system.-machine-learning-solutions

Machine learning : intelligence that learns by itself
https://www.bbva.com/en/machine-learning-intelligence-learns/

What Role Can Machine Learning And AI Play In Banking And Lending?
Breana Patel
https://www.forbes.com/sites/forbesfinancecouncil/2018/10/05/what-role-can-machine-learning-and-ai-play-in-banking-and-lending/#2dfc26cf4122

Machine learning analysis and modeling of interest rate curves
Mikhail Kanevski and Vadim Timonin

Volatility Forecasting using a Hybrid GJR-GARCH Neural Network Model
Soheil Almasi Monfared, David Enke

Réseaux de neuronnes
http://wikistat.fr

Data Mining Techniques and its Applications in Banking Sector
Dr. K. Chitra, B. Subashini

Technology for Planners
Trends, Spending, and the Rise of Robo Advisers
Roundtable moderated by Michael E. Kitces, CFP®, CLU®, ChFC®, RHU. REBC

Qu'est-ce que le machine learning ?
https://www.fimarkets.com/pages/machine-learning-finance.php

Le machine learning en finance
http://www.revue-banque.fr/banque-investissement-marches-gestion-actifs/article/machine-learning-en-finance

Appliqué à la finance, le deep learning modifie-t-il le cours de l'économie ?
https://www.sciencesetavenir.fr/high-tech/intelligence-artificielle/applique-a-la-finance-le-deep-learning-modifie-t-il-le-cours-de-l-economie_111364

7 Ways Fintechs Use Machine Learning to Outsmart the Competition
https://igniteoutsourcing.com/fintech/machine-learning-in-finance/

Stratégie automatique de trading en finance
http://www.xavierdupre.fr/app/ensae_teaching_cs/helpsphinx/specials/finance_autostrat.html

Le machine learning à l'épreuve de la réglementation
https://www.bankobserver-wavestone.com/machine-learning-a-lepreuve-de-reglementation/

Big data contre le blanchiment d'argent : machine learning, applications et réglementations financières
https://www.zdnet.fr/actualites/big-data-contre-le-blanchiment-d-argent-machine-learning-applications-et-reglementations-financieres-39852592.htm

Les regtechs au service des banques
https://fintech-mag.com/les-regtechs-au-service-des-banques/

Point de vue : Le risque de crédit pourrait-il être géré par une intelligence artificielle ?
https://www.groupe-estia.fr/point-de-vue-le-risque-de-credit-pourrait-il-etre-gere-par-une-intelligence-artificielle/

Les nouvelles méthodes de lutte contre la fraude bancaire en ligne
https://www.wavestone.com/app/uploads/2017/03/lutte-fraude-bancaire-en-ligne-nouvelles-methodes.pdf

Intelligence artificielle : enjeux pour le secteur financier
https://acpr.banque-france.fr/sites/default/files/medias/documents/2018_12_20_intelligence_artificielle_fr_0.pdf

IA, machine learning et relation client : la révolution n'en est qu'à ses débuts
https://www.relationclientmag.fr/Thematique/techno-ux-1256/Breves/machine-learning-relation-client-revolution-est-ses-debuts-333061.htm

Machine Learning for dummies
John Paul Mueller, Luca Massaron. Ed Wiley, 2016

Model risk managers eye benefits of machine learning
Louie Woodall
https://www.risk.net/risk-management/4646956/model-risk-managers-eye-benefits-of-machine-learning

How to Build a Product Recommendation System. Machine Learning Solutions
Michal Dyzma
https://www.netguru.com/blog/how-to-build-a-product-recommendation-system.-machine-learning-solutions

Banking Customer Segmentation and Machine Learning
Sachin Jahagirdar
https://www.linkedin.com/pulse/banking-customer-segmentation-machine-learning-sachin-jahagirdar

Machine Learning : intelligence that learns by itself
BBVA
https://www.bbva.com/en/machine-learning-intelligence-learns/

What Role Can Machine Learning And AI Play In Banking And Lending?
Breana Patel
https://www.forbes.com/sites/forbesfinancecouncil/2018/10/05/what-role-can-machine-learning-and-ai-play-in-banking-and-lending/#7aa530414122

Recommander des produits bancaires avec l'intelligence artificielle
Weave
https://weave.eu/competition-data-science-systeme-de-recommandation-bancaire/

Santander Product Recommendation Competition, 2nd Place Winner's Solution Write-Up
Tom Van de Wiele
http://blog.kaggle.com/2017/01/12/santander-product-recommendation-competition-2nd-place-winners-solution-write-up-tom-van-de-wiele/

Santander Product Recommendation
Ravi T
https://medium.com/@ravitee/santander-product-recommendation-ee4122d15072

Santander Product Recommendation. #1 Solution
idle_speculation
https://www.kaggle.com/c/santander-product-recommendation/discussion/26835#latest-549998

Machine Learning avec Scikit-Learn
Aurélien Géron. Dunod, 2017

Artificial intelligence and machine learning in financial services
Financial Stability Board, 2017
https://www.fsb.org/wp-content/uploads/P011117.pdf

AI and machine learning for risk management
Saqib Aziz, Michael Dowling
Rennes School of Business

How AI can help you get out of the "personal finance management" trap
Artashes Vardanian
https://hackernoon.com/how-ai-can-help-you-get-out-of-personal-finance-management-trap-f1dcfeed4431

Machine Learning Tutorial 5 - Big Data, Data Warehouse, Hadoop, Federation (Vidéo)
Caleb Curry
https://www.youtube.com/watch?v=bSgNB9bxSS0

Machine Learning in Banking Risk Management: a literature review
Martin Leo, Suneel Sharma, K. Maddulety

L'IA monte en puissance dans le secteur de la distribution française
https://www.lemagit.fr/actualites/252466444/LIA-monte-en-puissance-dans-le-secteur-de-la-distribution-francaise

Trois activités d'entreprise où l'intelligence artificielle présente déjà un réel impact sur la productivité
https://www.latribune.fr/opinions/tribunes/trois-activites-d-entreprise-ou-l-intelligence-artificielle-presente-de-ja-un-reel-impact-sur-la-productivite-782691.html

Big Value Data,Not Just Big Data!
https://www.fujitsu.com/global/Images/big-value-data-not-just-big-data.pdf

Le machine learning, une technologie qui veut du bien à l'industrie, selon Fujitsu
https://www.usine-digitale.fr/article/le-machine-learning-une-technologie-qui-veut-du-bien-a-l-industrie-selon-fujitsu.N397357

Intelligence artificielle : les banques françaises tâtonnent encore
https://www.latribune.fr/entreprises-finance/banques-finance/intelligence-artificielle-les-banques-francaises-tatonnent-encore-791045.html

OPCVM : quels sont les frais des fonds ?
https://www.cafedupatrimoine.com/archive/article/combien-coutent-opcvm

Les gestionnaires d'actifs français pèsent 4.000 milliards d'euros
https://www.lesechos.fr/2018/04/les-gestionnaires-dactifs-francais-pesent-4000-milliards-deuros-988257

Il en coûte un demi-million d'euros à une entreprise pour faire noter sa dette
https://www.latribune.fr/entreprises-finance/banques-finance/20130926trib000787085/il-en-coute-un-demi-million-d-euros-a-une-entreprise-pour-faire-noter-sa-dette.html

L'encours de dette mondiale franchit le cap des 100.000 milliards de dollars
https://www.lesechos.fr/2014/03/lencours-de-dette-mondiale-franchit-le-cap-des-100000-milliards-de-dollars-293746

La dette mondiale flambe à 250.000 milliards de dollars
https://www.lesechos.fr/finance-marches/marches-financiers/la-dette-mondiale-flambe-a-250000-milliards-de-dollars-1148484

Vidéo et automatisation, l'avenir de l'onboarding bancaire à distance ?
https://www.mindfintech.fr/article/16449/video-et-automatisation-l-avenir-de-l-onboarding-bancaire-a-distance/

Models and Méthods for Automated Credit Rating Prediction
Claude Gangolf

Machine learning et nouvelles sources de données pour le scoring de crédit
Christophe Hurlin, Christophe Pérignon, CAIRN
https://www.cairn.info/revue-d-economie-financiere-2019-3-page-21.htm

DONNER DU FUTUR AU TALENT

Fondé en 2008, Square est un cabinet de conseil en stratégie et organisation. 1er cabinet de conseil indépendant en France, en Belgique et au Luxembourg, Square est, avec ses 700 consultants, l'un des rares acteurs du marché à proposer une gamme d'expertises aussi étendue.

Square guide ses clients en mettant à leur disposition ses compétences et son expérience sur 8 domaines d'excellence :

INNOVATION
Square accompagne ses clients dans la transformation de leur dynamique d'innovation. Nos consultants, par leur approche sur-mesure, aident à concevoir, industrialiser et gouverner l'innovation pour assurer la croissance durable des entreprises et leur transformation en entité socialement et écologiquement responsable.

PEOPLE & CHANGE
Square aide ses clients à acquérir, fédérer et développer le capital humain de leur organisation. Afin de créer davantage d'engagement au sein des équipes, nos interventions portent principalement sur l'adaptation des méthodes de travail aux changements opérationnels et culturels, l'efficacité des directions des ressources humaines et le développement des compétences.

MARKETING
Square accompagne ses clients sur l'ensemble du spectre marketing : marketing stratégique, marketing relationnel, marketing de l'offre, communication, tarification, satisfaction clients. Nos expertises, initialement centrées sur les secteurs de la banque et de l'assurance, s'adressent désormais à l'ensemble des industries ou services B2C.

DATA
Square élabore des stratégies Data et assure leurs déclinaisons opérationnelles à travers la conduite de projets de Data Management, Data Analyse et Data Science. Notre approche experte et pragmatique vise à valoriser et sécuriser le patrimoine de données des entreprises.

DIGITAL
Square accompagne ses clients dans l'élaboration de leur stratégie digitale, la conception et la mise en œuvre de nouveaux parcours digitaux pour leurs clients ou leurs collaborateurs, ainsi que dans l'ensemble des chantiers d'acculturation interne et d'accompagnement aux nouvelles méthodes de conception.

RISK & FINANCE
Square prend en charge le pilotage des programmes de maîtrise des risques financiers et non financiers, ainsi que la transformation des fonctions Risque et Finance face à l'évolution des dispositifs prudentiels et à l'irruption des problématiques liées à la maîtrise de la donnée.

REGULATORY & COMPLIANCE
Square conseille ses clients dans le déploiement des nouvelles réglementations, ainsi que dans l'optimisation et le renforcement des dispositifs de contrôle. Ce domaine d'excellence s'appuie sur une communauté d'experts de 130 consultants qui, outre les missions auprès des clients, conduit d'importants travaux d'investigation et de publication.

SUPPLY-CHAIN
Square assure l'excellence opérationnelle de la logistique, des achats aux derniers kilomètres, avec des parcours clients différenciants. Nos experts conçoivent des solutions omnicanales mettant en œuvre les meilleures pratiques des systèmes d'informations, de la mécanisation à la robotisation.

Rédigé par les consultants Square des domaines d'excellence Data et Risk & Finance, ce focus propose de revenir sur une tendance forte au sein des programmes de digitalisation des services financiers : le recours accru au Machine Learning. Si le recours aux mathématiques ne constitue vraiment pas une innovation, l'ampleur de ce déploiement et la transversalité des usages possibles ouvrent un champ des possibles tout à fait nouveau. Afin de démystifier la mise en oeuvre de cette forme d'Intelligence Artificielle, ce focus propose donc, après une brève présentation du contexte favorable à l'explosion du Machine Learning, de revenir sur plusieurs cas d'usages opérationnels en banque et ainsi illustrer les défis qu'il reste à relever.

Crédits photos - Unsplash : Taylor Vick (pp. 1, 6), Imgix (p.8), Andras Vas (p. 14, 29), Blake Connally (p. 30, 39), Christian Wiediger (p. 40, 42).

CONTACTS

JULIEN GUIBERT
PARTNER
+33 6 67 56 90 02
julien.guibert@square-management.com

MARC CAMPI
PARTNER SQUARE
+33 6 84 02 68 59
marc.campi@square-management.com

ADRIEN AUBERT
ASSOCIATE PARTNER
+33 6 69 63 06 01
adrien.aubert@square-management.com

Square

Square

DONNER DU FUTUR AU TALENT

square-management.com