



CABINET ROUX
Exercice 2020

Projets de Recherche et Développement
(Crédit d'impôt en faveur de la recherche)

SOMMAIRE

1. PRÉSENTATION DE L'ENTREPRISE	4
1 L'entreprise.....	4
2 Gestion de l'innovation	4
2. SYNTHÈSES TECHNIQUES	6
1 TRANSFORMATION DATA ALGORITHMIQUE.....	6
1.1 Objectifs du projet	6
1.2 État de l'art	7
1.3 Aléas, incertitudes scientifiques, verrous technologiques	10
1.4 Travaux R&D réalisés, démarche expérimentale	11
B. Indicateurs de R&D	15
C. Conclusions	15

PRÉSENTATION DE L'ENTREPRISE

1 L'entreprise

Cabinet Roux, créé en 1888, est une société d'expertise et de conseil spécialisée dans l'étude et la réalisation de missions d'évaluation, d'analyse et de gestion des :

- Patrimoines corporels immobiliers ;
- Matériels des entreprises industrielles ;
- Grands investisseurs ;
- Opérateurs de l'état ;
- Collectivités publiques et territoriales ;
- L'accompagnement des assurés avant et après sinistre.

Notre objectif principal est d'offrir un savoir-faire dans tous les domaines de l'expertise et de valeurs :

- Expertise préalable bâtiment, matériel, pertes d'exploitation ;
- Risques Industriels ;
- Valorisation des Patrimoines privés, particuliers et professionnels en valeur de marché et valeur d'assurance ;
- Audit technique (environnement, risque, prévention) ;
- Expertise après sinistre pour le compte des assurés (industriels et commerciaux, auprès des TPE, PME, ETI, Grands Comptes et particuliers).

Cabinet Roux est présent sur tous les pôles économiques majeurs sur le plan mondial et offre à ses clients, industriels et commerciaux, les moyens de préserver le caractère normatif et homogène des études et des valeurs. Par ailleurs, Cabinet Roux est l'un des membres fondateurs d'[UPEMEIC](#), et signataire de sa charte déontologique.

2 Gestion de l'innovation

Notre équipe R&D est placée sous la direction de M. Benoît Cordesse. Nous investissons aujourd'hui dans le développement et la mise en œuvre d'outils d'aide à la décision pour la classification des biens et la valorisation automatique en fonction des capacités des biens expertisés.

Le choix des algorithmes pertinents, leur optimisation et les améliorations nécessaires à leur mise en œuvre sont des sujets au cœur du troisième pilier de la transformation numérique, qu'est le Big-Data et l'exploitation des données.

Pour notre recherche en 2019, nous avons collaboré avec les cabinets Operation Data (AREMUS CREATION ET DEVELOPPEMENT) et Frédéric Lefebvre-Naré (ISEE). Spécialisés en Data Science et en Intelligence Artificielle ils ont su ajouter leur expertise à

la nôtre pour transformer le secteur et développer un outil de classification des biens et la valorisation automatique en fonction des capacités des biens expertisés.

SYNTHÈSES TECHNIQUES

Chef de projet : M. Benoît Cordesse

1 TRANSFORMATION DATA ALGORITHMIQUE

1.1 Objectifs du projet

Le secteur des évaluations d'actifs industriels et de l'expertise assurance est un secteur concurrentiel, dont les marchés sont nationaux mais dont la pérennité est incertaine. En effet, les technologies d'évaluations, qui sont _elles_ globales, menacent de plus en plus d'entrer dans le marché.

Cabinet Roux se considère comme leader technologique en France pour l'évaluation d'actifs industriels et l'expertise assurance. Ainsi, il veut tirer profit de ces outils technologique pour garder sa position de leader et permettre d'atteindre ses objectifs : 60% de croissance dans les 5 prochaines années.

Investir dans l'automatisation data-IA est donc un point stratégique pour accomplir cette ambition.

Cet investissement permettrait de :

- Outiller le processus d'expertise pour gagner en efficience (commerciale et technique) donc en marge nette (potentiel >1M€/an)
- Rendre apprenante l'expertise (capitalisable, cherchable) pour outiller (ci-dessus) et gagner en vitesse de croissance
- Devenir incontournable en évaluation d'actifs industriels/immobiliers. Diversification possible valorisant les data (prévention, support courtage...)

Un audit interne de notre niveau de maturité technologique a été effectué par OperationData. Les résultats, disponible ci-dessous, soulignent notre crédibilité dans le secteur de l'automatisation data-IA et les opportunités du Cabinet Roux.

Maturité data-algorithmique absolue	↑
	Très élevée - Un datalake interconnecté, une boucle algorithmique d'optimisation online des produits & services et/ou une partie importante du business-model directement issu de la valorisation de data.
	Elevée - Un SI / datalake interconnecté, outils de simulation/optimisation et des ressources prédictives dédiées (équipe interne et/ou logiciels ayant passé la preuve de concept), mode SaaS.
	Moyenne* - Des outils fonctionnels de type CRM, ERP, ... utilisés et interconnectés, outils de simulation numérique mais sans encore de data analytics dédiés.
	Faible - Peu de digitalisation des processus, sauf outils fonctionnels de type CRM, ERP, comptabilité dématérialisée, page web et réseaux sociaux etc., mais non interconnectés.
	Très faible - Pas de digitalisation significative des processus, pas d'outils fonctionnels (hors bureautique).

Après une analyse de la maturité data-algorithmique concurrentielle, nous considérons que nous disposons déjà, de par nos moyens, d'une avance concurrentielle en France, mais que nos concurrents étrangers (aux Etats-Unis notamment) disposent d'une avance supplémentaire. Ainsi, il est important pour le

Cabinet Roux d'investir dès maintenant dans ce secteur pour pouvoir rester concurrentiel face aux cabinets Américains.

C'est pour suivre cet axe de développement stratégique qu'en 2019 le Cabinet Roux s'est fixé comme mission de développer des algorithmes d'apprentissage automatique pour harmoniser au mieux les pratiques d'estimation des biens.

Dans la continuité, les travaux de R&D de l'année 2020 se sont principalement concentrés sur la création d'un outil d'aide à la décision (appelé Predict Value) simple qui permettrait, en renseignant différentes catégories d'informations (pays, la taille de l'usine, les caractéristiques des machines) d'obtenir une première estimation de valeur (savoir à quel montant assurer). C'est le principal but recherché de Predict Value.

L'autre objectif est d'estimer le coût des sinistres potentiels. Savoir à quel risque maximum un site, notamment industriel, pourrait être exposé. C'est ce qu'on appelle le SMP (Sinistre Maximum Possible). Il correspond au montant des dommages matériels le plus important pouvant résulter d'un événement.

Le dernier objectif est de déterminer la PE (Perte d'exploitation).

A travers Predict Value, Cabinet Roux veut aider le département assurances de Sanofi à montrer que l'entreprise a une bonne gestion de son patrimoine. Cabinet Roux pourra également démontrer sa capacité à adopter une approche "produit" et "data", innovant de cette manière sur son cœur de métier. En effet, Predict Value sera le premier produit hébergé en externe dans l'histoire du cabinet.

1.2 État de l'art

1.4.1. Le digital et l'Intelligence Artificielle dans l'estimation des capitaux mobiliers et immobiliers

L'Intelligence Artificielle et la Data Science sont des domaines qui bénéficient depuis quelque temps d'une attention sans précédent et séduisent de plus en plus de secteurs. Véritable opportunité pour se décharger de certaines tâches, gagner en productivité, augmenter sa puissance de calcul ou analyser des données intéressantes. Néanmoins, elle représente aussi une menace pour les entreprises et secteurs traditionnels, qui ne sont pas structurés pour l'adopter et qui voient une nouvelle forme de concurrence arriver sur les marchés [1].

Dans le domaine de l'expertise immobilière en France, le digital prend la forme de site internet ou de bases de données. Néanmoins, avec l'arrivée de concurrents digitaux en ligne pour les ventes aux particuliers, le secteur subit des transformations importantes et s'oriente de plus en plus vers le digital [2]. L'intelligence artificielle n'est pas encore

stablement implémentée dans le secteur mais l'opportunité est réelle et beaucoup étudiée dans le cadre de la spéculation immobilière [3].

Dans la gestion d'actif, domaine directement lié à l'évaluation d'actifs, la Data Science et l'Intelligence Artificielle sont de plus en plus utilisées et ont prouvés leur utilité. Dans la maintenance prédictive d'ensemble d'actifs, par exemple, le digital et la Data Science permettent la Predictive Maintenance (PdM – Maintenance prédictive). Cette fonction permet de réduire les coûts opérationnels, d'augmenter la productivité et la sûreté. Implémentée de l'Intelligence Artificielle (Machine Learning) les programmes de maintenance prédictive se sont considérablement améliorés et répandus [4].

De même, les stratégies d'investissements sont maintenant évaluées par l'intelligence artificielle, utilisant la Valeur Actuelle Nette (VAN) des biens et composants [5]. Les stratégies d'investissement mobiliers et immobiliers sont donc optimisées par le biais d'intelligence artificielle et de Data Science.

Bien qu'il n'existe pas actuellement de programme d'évaluation du capital mobilier, des assurances de particulier, comme la GMF, proposent un outil d'évaluation du capital mobilier dans lequel il faut rentrer manuellement les valeurs des biens dans chaque pièce [6].

Les travaux du Cabinet Roux ne sont donc pas adressés par la littérature mais s'inscrivent dans la continuité de celle-ci et des tendances du secteur.

1.4.2. La classification et évaluation automatisée de biens

La classification automatisée supervisée ou non est un domaine bien renseigné dans la littérature scientifique [7][8]. Les méthodes de classement ayant pour objet d'identifier la classe d'appartenance d'objets définis par leur description [7].

Quant à la classification d'une liste de mots en particulier, la recherche accorde ces dernières années, beaucoup d'importance au traitement des données textuelles [8]. Ces données, obtenues par des questionnaires ou encore les réseaux sociaux sont une des sources principales des données alternatives aujourd'hui. Ainsi le domaine de la fouille de texte (text mining) s'est développé pour répondre à volonté à la gestion par contenu des sources volumineuses de textes [8].

Néanmoins, il n'existe pas de littérature sur une évaluation automatisée de liste de biens industriels. La valeur des biens dépendant de nombreux facteurs, notamment leur année d'acquisition, mais aussi le secteur dans lequel ils sont utilisés. On peut trouver de la donnée à propos de méthodes de calculs d'experts pour estimer un bien

industriel, qui est donc théorisé, mais qui nécessite actuellement l'intervention d'un humain pour chaque calcul et n'est donc pas automatisée [9].

Nous avons donc été confrontés à l'absence d'articles scientifiques présentant des travaux de classification et de valorisation basés sur des algorithmes d'apprentissage automatique particulièrement efficaces pour de telles activités modernes. Les méthodes existantes nécessitent l'interaction humaine ce qui peut engendrer des erreurs car la ligne de production ne serait par exemple complète. En effet, les données en termes de prix marché sont généralement obtenues auprès des professionnels revendeurs ou bien d'entreprises ayant une bonne connaissance du bien à évaluer. Dans ce cas, il est nécessaire d'appeler des représentants et, si possible, d'aller à une vente aux enchères et poser beaucoup de questions. Ceci nécessite également de contacter les fabricants de machines, les vendeurs d'occasion/neuf, consulter les revues spécialisées, les journaux, s'informer auprès de partenaires ou entreprises similaires, ...

C'est dans cette partie de la littérature que cette recherche s'inclus, en tentant de s'adresser au manque d'algorithme évaluant automatiquement les biens d'un industriel (mobiliers et immobiliers).

Nos outils permettent donc de remédier à toutes ces faiblesses en automatisant la classification et la valorisation des biens. Effectivement, Predict Value va accompagner la transformation du métier de l'expertise préalable, et aider Cabinet Roux à se démarquer de ses concurrents (Expertise Galtier par exemple). En apportant à ses clients des éléments de business intelligence, à travers des produits ergonomiques et simples d'utilisation, Cabinet Roux est acteur de la transformation de son cœur de métier.

1.4.3. Bibliographie et sitographie

[1] Portnoff André-Yves, Soupizet Jean-François, « Intelligence artificielle : opportunités et risques », *Futuribles*, 2018/5 (N° 426), p. 5-26. DOI : 10.3917/futur.426.0005. URL : <https://www.cairn-int.info/revue-futuribles-2018-5-page-5.htm>

[2] Gardès, Nathalie. « Digitalisation du secteur immobilier : la proposition de valeur phygitale au cœur de la performance », *La Revue des Sciences de Gestion*, vol. 299-300, no. 5, 2019, pp. 133-146.

[3] Mihnea Constantinescu, « Machine-Learning Real Estate Valuation: Not Only a Data Affair », 2019, A Medium publication sharing concepts, ideas, and codes. <https://towardsdatascience.com/machine-learning-real-estate-valuation-not-only-a-data-affair-99d36c92d263>

[4] Aremu, Oluseun & Salvador Palau, Adrià & Parlikad, Ajith Kumar & Hyland-Wood, David & Mcaree, Peter, « Structuring Data for Intelligent Predictive Maintenance in Asset Management. », 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved. FAC PapersOnLine 51-11 (2018) 514-519

[5] Jeanne DEMGNE, « MODELISATION D'ACTIFS INDUSTRIELS POUR L'OPTIMISATION ROBUSTE DES STRATEGIES DE MAINTENANCE », 2015, UNIVERSITE DE PAU ET DES PAYS DE L'ADOUR

[6] https://services.gmf.fr/services/aides/outilsaide/capital_mobilier.html

[7] Yves Lechevallier, « cours Méthodes de classification supervisée: les méthodes de segmentation ou les arbres de décision » [archive], Master ISI, Université Paris-Dauphine et Agro ParisTech. https://www.lri.fr/~antoine/Courses/Master-ISI/ISI_arbre.pdf

[8] yasmine Hanane zeggane Mokhtar, « Algorithmes d'apprentissage pour la classification de documents » 2009, Université de Mostaganème, Algérie
https://www.memoireonline.com/12/09/2917/m_Algorithmes-dapprentissage-pour-la-classification-de-documents0.html

[9] https://machinelivepartner.com/wp-content/uploads/2016/02/methodes_estimation-MLP.pdf

1.3 Aléas, incertitudes scientifiques, verrous technologiques

Lors de la conception de notre outil Predict Value, nous avons été confrontés aux incertitudes suivantes :

D'une part, notre outil Predict Value doit servir notre expertise et notre neutralité. Predict Value est vu comme un produit d'appel pour pouvoir faire passer une expertise humaine en aval de Predict Value pour raffiner l'estimation fournie. Pourrions-nous donc sortir un outil d'aide à la décision et créer un cas d'usage visant à valider la capacité prédictive de notre expertise et la fiabilité de notre modèle de données ?

D'autre part, notre outil doit être juste en termes d'estimation de valeur. Nous devons ainsi obtenir des hypothèses fiables pour déterminer la valeur de reconstruction à l'identique d'un site existant. Ainsi, pourrions-nous assurer une fiabilité des données pour des estimations justes ?

De plus, Predict Value doit être pédagogique et didactique pour communiquer auprès d'autres utilisateurs l'expertise du dommage. Comment pourrions-nous assurer ceci ?

Enfin, nous devons réussir à tenir le jeu d'équilibriste de la conception de Predict Value, un outil à court terme pour Sanofi, tout en gardant en tête que demain ce produit pourrait / devrait être présentée et vendu à d'autres prospects de Cabinet

Roux (clients grands comptes existants, mais aussi des cibles très différentes qui pourraient être intéressées, telles que les courtiers en assurance).

1.4 Travaux R&D réalisés, démarche expérimentale

En 2020, nos travaux de R&D ont porté principalement sur les axes de développement suivants :

- Valorisation automatique des biens
- Estimation des coûts des sinistres potentiels
- Détermination de la perte d'exploitation

1.4.1. Valorisation automatique des biens

.1.4.1.1 Optimisation de Predict Value

Nous avons développé en partenariat avec Océane Cordesse (Data Analyst Freelance) les algorithmes Explore et Predict VALUE pour notre client Sanofi dans le but de gérer leur patrimoine industriel.

L'algorithme Explore est un outil de pilotage qui offre une exploitation intuitive des données. Predict Value permet quant à lui de prédire la valeur d'un site industriel pharmaceutique à partir de quelques données techniques. La description de Predict Value est détaillée dans l'annexe I.

Afin d'améliorer notre algorithme Predict Value, nous avons mis en place un partenariat avec l'ICAM. Deux algorithmes ont été étudiés : l'algorithme structuration et l'algorithme testeur.

Ces algorithmes ont pour rôle d'estimer la valeur unitaire des équipements principaux présents dans la base de données pour donner une estimation de la valeur à neuf d'un site industriel partie équipement uniquement. Au cours de ce projet nous avons été amenés à les modifier partiellement surtout les fonctions d'optimisation.

Dans le cas de l'algorithme structuration, nous avons ajouté la possibilité de modifier les poids en fonction des « main process » comme pour « Supply Chain » pour le client Sanofi. Le deuxième ajout était de créer une liste et une table « table_POIDS_CUMULE » pour récupérer les poids du main process et du kind of process et les transférer à l'algorithme testeur. On vient alors enregistrer et exporter cette table comme toutes les autres, elle servira également de fichier source pour l'algorithme testeur.

En ce qui concerne l'algorithme testeur, plusieurs modifications ont été apportées comme la création de différentes boucles internes et l'ajout de la division des *kind of process* par leur(s) parent(s). L'ensemble des modifications réalisées est détaillé dans l'annexe II.

Commenté [MM1]: Pourriez-vous préciser en quoi consistait l'amélioration apportée ?

Après l'optimisation des algorithmes, nous nous sommes intéressés à l'optimisation de la prédiction. Pour faire, nous avons divisé le travail en 5 étapes :

- Changer le « poids »
- Modifier la variable nommée "Maxeval" qui représente le nombre d'itération des formules employés
- Changer les deux formules d'optimisation (nlm et nloptr)
- Modifier les paramètres de la formule nloptr
- Changer la formule d'estimation

Concernant la première modification : le changement du « poids », nous avons essayé plusieurs tests. Nous avons dans un premier temps, augmenté le poids par palier (80%/85%/90%), cela a eu pour impacte d'améliorer la performance de l'algorithme. Cependant, un poids trop élevé nécessite un nombre de données trop conséquent à apporter. Au contraire, diminuer le poids diminue les performances de l'algorithme. Nous avons décidé de garder un poids assez haut pour maximiser la performance sans que cela nécessite une trop grande quantité de données à apporter. Ci-dessous un tableau récapitulatif nos différentes tentatives et leurs résultats :

Tableau 1 : Tableau récapitulatif des différentes tentatives testées

MAIN_PROCESS	75% - 75% - 2 000			85% - 80%			85% - 85%		
	Erreur	KOP	BIEN	Erreur	KOP	BIEN	Erreur	KOP	BIEN
Biologics	41.78 %	6	24	36.89 %	9	33	36.00 %	9	34
Chemical process	37.82 %	4	20	36.67 %	8	32	36.00 %	8	34
Injectables	33.40 %	9	38	28.55 %	15	65	26.64 %	15	71
Pharmaceutical process	18.47 %	11	40	18.12 %	20	85	18.72 %	20	93
R&D	31.43 %	4	110	24.13 %	7	145	26.93 %	7	184
Supply Chain	103.00 %	5	22	55.38 %	7	29	54.75 %	7	31
Vaccines	28.80 %	17	56	37.67 %	33	111	36.67 %	33	121

75/85% - 75% 6 BIENS max			75/85% - 75% pas limite BIENS		
Erreur	KOP	BIEN	Erreur	KOP	BIEN
35.56 %	6	27	35.56 %	6	27
27.25 %	4	24	27.25 %	4	24
33.10 %	9	40	33.10 %	9	40
18.22 %	11	44	18.22 %	11	44
31.43 %	4	110	31.43 %	4	110
100.50 %	5	25	57.12 %	7	29
31.56 %	17	63	31.56 %	17	63

Nous avons donc validé un poids de 85% pour le main process et 75% pour les kind of process sans limiter le nombre de kind of process dans le calcul.

Pour la deuxième stratégie : Modifier la variable « Maxeval » qui représente le nombre d'itération, nous avons réalisé des tests en l'augmentant et en le diminuant. Les résultats n'ont pas été concluants, nous avons donc gardé les paramètres de base.

Ensuite, nos travaux ont porté sur le changement des formules d'optimisation et des paramètres de la formule. Les différents travaux ont montré que les changements effectués n'ont pas d'impact sur la performance de l'algorithme si sur la durée d'exécution.

Enfin, nous nous sommes intéressés au changement de la formule d'estimation qui lie la capacité d'un équipement à son prix. Après l'analyse des données, nous nous sommes rendu compte que l'évaluation de notre modification a été impossible.

La dernière fonctionnalité optimisée a été le passage d'un client à un autre, pour cela, nous avons pu créer un fichier Excel automatisé capables de créer les fichiers sources de l'algorithme Structuration avec le moins d'actions manuelles possible. Ce fichier permet au client de gagner du temps lors qu'il veut analyser un nouveau client.

Ces travaux de recherche ont permis d'avoir un algorithme Predict Value amélioré en termes de performance et de temps d'exécution.

.1.4.1.2 Méthodologie de calcul de la valorisation

Dans cette partie nous présentons la méthodologie de prédiction de valeur d'une usine SANOFI.

A) Définition de l'arborescence

Pour estimer la valeur mobilière des biens dans tout type d'usine, il nous faut comprendre les biens nécessaires selon les types d'usines. En effet, une usine est constituée de plusieurs niveaux de description arborescente du patrimoine physique : une ENTREPRISE est constituée d'un ÉTABLISSEMENT (sites), chaque établissement peut comprendre plusieurs BÂTIMENTS (dont la surface et la valeur immobilière figurent généralement dans AppValue). Chaque bâtiment comprend un ou plusieurs SECTEURS (correspondant chacun à un type d'activité physique : par exemple production, entreposage, bureaux... dans un même bâtiment) ; enfin, chaque secteur comprend une liste de BIENS.

Nous avons donc choisi de décrire la relation arborescente par une « hiérarchie parents-enfants ». Plus précisément, nous avons structuré les données comme suit : Le type de bâtiment (« Brasserie ») est appelé grand-parent, le type d'unité (« Bureaux », « Laboratoire » ...) est un parent, enfin, le type de bien (« Balance ») sera catégorisé en enfant.

B) Méthodologie de calcul de la valorisation

Dans le but de valoriser automatiquement les biens, nous avons défini les trois points clés suivants :

- Périètre de calcul : Nous définissons le périmètre de calcul comme un ensemble de lignes et de champs de caractérisation, un ensemble de parents et enfants et un nombre de bâtiment.
- Méthodologie de calcul : Pour évaluer la performance de notre modèle de machine learning, nous avons séparé notre jeu de données en deux parties distinctes :
 - Le training sert (80% du périmètre), qui va nous permettre d'entraîner notre modèle

- Le test set (20% du périmètre), qui nous permet de mesurer l'erreur du modèle / la marge d'erreur
- Modèle algorithmique : Nous utilisons l'algorithme **XGBoost** plébiscité par la communauté data science pour sa vitesse d'exécution et sa performance.

Commenté [MM2]: Pourquoi vous avez utilisé cet algorithme et non pas Predict Value?

Notre processus algorithmique de valorisation en fonction des actions utilisateurs est illustré à la Figure 1 ci-dessous.

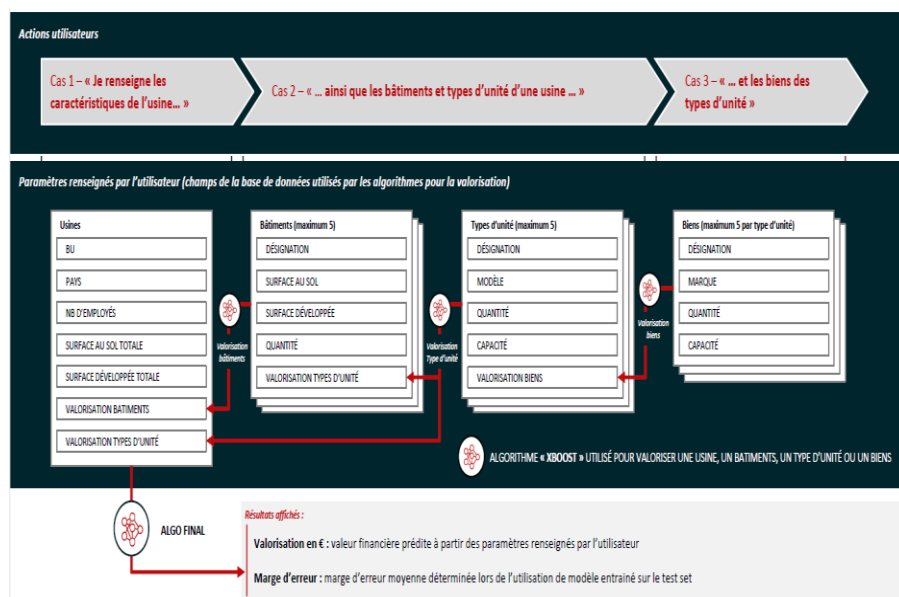


Figure 1 : Processus algorithmique de valorisation en fonction des actions utilisateurs

Les résultats affichés sont alors la valeur financière prédite à partir des paramètres renseignés par l'utilisateur avec une marge d'erreur.

1.4.2. Estimation des risques (sinistre maximal possible)

L'objectif de ces travaux est de créer un « aspirateur » récupérant sur internet et les sites d'information des données sur les sinistres et/ou incendies.

Nous avons alors mis en place un algorithme pour récupérer automatiquement les articles et un outil à destination des commerciaux pour « qualifier » l'article : Pas un sinistre, Sinistre mais pas mon secteur et sinistre.

Nous avons utilisé des algorithmes deep learning pour qualifier automatiquement les articles.

Les résultats obtenus par l'aspirateur ont été comparés avec les données de notre prestataire ATCOM. D'après les résultats d'analyse, notre aspirateur serait

Commenté [MM3]: Pourriez-vous nous communiquer la documentation technique décrivant les grandes lignes des travaux menés dans cette partie?

Commenté [MM4]: S'agit-il d'un algorithme que vous avez développé ? adapté ? Pourriez-vous nous communiquer les travaux réalisés et les difficultés rencontrées pour mettre en place cet algorithme ?

Commenté [MM5]: Pourriez-vous donner plus d'informations concernant ces algorithmes ?

théoriquement capable de remplacer totalement ATCOM sur ces signalements en communs. Cependant, il ne serait pas plus rapide que ATCOM.

Afin d'étendre le flux, il faudrait relier notre aspirateur à de nouvelles sources. Néanmoins, nous n'avons pas encore connaissance de l'intégralité de ces dernières. Nous envisageons donc de continuer à étudier les signalements aspirateurs étant arrivés après ATCOM, et de réaliser des recherches afin de trouver de nouvelles sources.

Cette solution implique l'envoi de plus de signalements à nos chargés de clientèle. Il est donc nécessaire d'améliorer notre filtre « détecteur de sinistres » pour tenter de réduire les faux-positifs.

1.4.3. Détermination de perte d'exploitation

Commenté [MM6]: Pourriez-vous nous communiquer la documentation technique décrivant les grandes lignes des travaux menés dans cette partie?

B. Indicateurs de R&D

Les travaux de recherche et développement présentés dans cette synthèse repose sur notre expertise du métier de l'évaluation mobilière et immobilière mais aussi de développement informatique. En effet, les ingénieurs qui ont participé aux développements justifient d'un haut niveau académique et d'une connaissance approfondie de l'évaluation des sites industriels et de l'informatique.

De plus, ces travaux ont été réalisés en collaboration avec l'Institut Catholique d'Arts et Métiers (ICAM).

Commenté [MM7]: Est-ce qu'il y a eu d'autres collaborations en 2020 ?

Les objectifs que nous nous étions fixés en début de projet n'étaient pas adressés par l'état de l'art. Nos compétences en évaluation de biens industriels et les connaissances de nos collaborateurs en développement de solutions Big Data et Intelligence Artificielle et notre maîtrise de la technologie nous ont permis de repousser les frontières de son utilisation en concevant de nouvelles solutions complexes, en phase avec notre ambition.

C. Conclusions

L'objectif de nos travaux de recherche de l'année 2020 était de concevoir un outil d'aide à la décision (Predict Value) simple qui permettrait, en renseignant différentes catégories d'informations (pays, la taille de l'usine, les caractéristiques des machines) d'obtenir une première estimation de valeur (savoir à quel montant assurer). A travers Predict Value, Cabinet Roux veut aider le département assurances de Sanofi à montrer que l'entreprise a une bonne gestion de son patrimoine. Cabinet Roux pourra

également démontrer sa capacité à adopter une approche “produit” et “data”, innovant de cette manière sur son cœur de métier. En effet, Predict Value sera le premier produit hébergé en externe dans l’histoire du cabinet.

L’autre objectif est d’estimer le coût des sinistres potentiels. Savoir à quel risque maximum un site, notamment industriel, pourrait être exposé.

Le dernier objectif est de déterminer la perte d’exploitation.

En apportant à ses clients des éléments de business intelligence, à travers des produits ergonomiques et simples d’utilisation, Cabinet Roux est acteur de la transformation de son cœur de métier.

AXYS Consultants

Exercice 2020

Projets d'Innovation

(Crédit d'Impôt en faveur de l'Innovation)

SOMMAIRE

3. PRÉSENTATION DE L'ENTREPRISE	20
1 L'entreprise.....	20
2 Gestion de l'innovation	20
4. SYNTHÈSE TECHNIQUE.....	21
1 Développement d'outils technologiques spécialisés.....	21
1.1 Objets du projet	21
1.2 Contexte économique du projet.....	22
1.3 Description des travaux effectués.....	23
1.4 Conclusions.....	33
5. ANNEXES	35

PRÉSENTATION DE L'ENTREPRISE

1 L'entreprise

Axys Consultants, société de conseil en management fondée en 1987 et comptant 163 consultants, accompagne ses clients, grands groupes et ETI, sur leurs problématiques de transformation dans les domaines des achats, de la finance, du marketing digital et du commerce.

Axys Consultant place l'usage de la donnée comme catalyseur de la conduite du changement notamment à travers ses deux BU transverses, Data et Change.

En effet, la mutation actuelle du marché du conseil et l'émergence de nouveaux modèles disruptifs et l'accroissement de problématiques basés sur l'IA, la Blockchain et la RPA, transformant la chaîne de valeur, incitent les cabinets à repenser la manière dont ils conseillent leurs clients.

Afin de se positionner en tant que leader sur ces problématiques, la société a ouvert son LAB IA début 2019. L'objectif est de démontrer les avantages offerts par le Machine Learning pour ainsi accroître la valeur ajoutée de ses prestations grâce à l'IA « As a Service ».

De cette manière, les travaux de recherche menés sur ces sujets nous permettent d'intégrer des prestations dans domaines à forte valeur ajoutée à notre offre de services, tout en ayant la capacité de doter nos consultants de nouveaux outils dans le but d'améliorer notre productivité et la qualité de nos prestations.

2 Gestion de l'innovation

Afin de relever ces défis, le LAB IA d'Axys Consultants peut s'appuyer sur l'expertise et l'expérience de ses équipes. En effet, celle-ci est constitué d'ingénieurs développeurs, de Data Scientists et Data Analysts bénéficiant tous d'une expérience significative dans le secteur de la tech.

De plus, nos travaux font régulièrement l'objet de publications sous forme de tribunes ou de livres blanc sur notre page internet ainsi que sur des sites spécialisés.

Par ailleurs, des travaux d'innovation sont également menés par les équipes Axys Consultants qui travaillent notamment au développement de nouvelles solutions technologiques à destination de ses clients.

SYNTHÈSE TECHNIQUE

Chefs de projet : **M. Jean-Luc Marini**
M. Lionel Bianchi
M. Marc Mironneau
M. Julien Samarcq

1 Développement d'outils technologiques spécialisés

1.1 Objets du projet

Durant l'année 2020, Axys Consultant a travaillé sur diverses solutions technologiques afin de proposer des solutions innovantes et adaptées aux besoins de ses clients exerçant dans de multiples secteurs d'activités.

Dans un premier temps, nous avons collaboré avec notre client **BCA Expertise**, une entreprise d'expertise automobile qui intervient sur toute la nature des dommages qu'un véhicule a subi lors d'un accident. Nous avons développé une application Blockchain permettant d'optimiser le processus de gestion des véhicules gravement accidentés. Cette plateforme permet aux différentes parties prenantes de collaborer plus efficacement via un ordonnancement des actions et une identification des acteurs concernés à chacune des étapes tout en consignait l'ensemble des décisions et des actions effectuées au cours du processus.

En outre, nous étions intervenus dans l'entité IES de la société BCA qui opère plusieurs micro-services internes basés sur l'Intelligence Artificielle dans le domaine de l'assurance automobile notamment, les services de sécurité, supervision, gestion des clients, stockage des données, Backup, service Mesh, développements IA ainsi que le développement des logiciels.

Si aujourd'hui les solutions IA de l'IES rendent les services attendus, ils n'offrent pas les meilleures garanties en termes de scalabilité, de gestion multi-clients, de supervision et de bonnes pratiques de déploiement. De plus, l'interaction entre les différents services n'est pas conçue de façon optimale. Nous avons donc développé une architecture Micro-Services SaaS dont l'objectif est de supporter la charge induite par le nouveau positionnement de l'IES en tant que fournisseur de services, permettant d'intégrer rapidement de nouvelles offres de services et fournir tous les outils nécessaires au suivi de la production, tout en limitant au maximum l'impact de ce changement d'architecture sur les micro-services eux-mêmes.

De plus, nous avons collaboré avec l'opérateur télécom **La Poste Mobile** qui a souhaité développer un algorithme multicritère de gestion de la fragilité financière permettant d'identifier et de quantifier le risque de défaut de paiement des clients lors de la

souscription pour un opérateur télécom. L'objectif est ainsi de diviser par deux le taux de résiliation des abonnements à trois mois tout en limitant les demandes de garanties aux clients les plus à risque.

Nous avons enfin développé, pour notre client **SKODA**, un outil de prédiction et pilotage des leads permettant de prédire le nombre de leads qualifiés dans le secteur automobile en fonction du plan marketing (publicité, messages, budget par partenaire, etc.,) ainsi que des données externes.

Les travaux menés en 2020 s'inscrivent dans la continuité des initiatives de recherche entreprises depuis le lancement du projet dans le but d'offrir une plateforme toujours plus innovante et performante à nos utilisateurs.

1.2 Contexte économique du projet

Aujourd'hui, de plus en plus d'entreprises fournissant des solutions blockchain dans les différents secteurs d'activités voient le jour. IBM constitue un acteur majeur dans ce domaine fournissant différentes solutions d'approvisionnement en blockchain notamment, les solution Trust Food et Trade Lens qui sont deux projets de très grande envergure réussis par IBM.

Trust Food est un réseau mettant en relation les acteurs de l'approvisionnement alimentaire par le biais d'un enregistrement partagé et permissionné des données du système alimentaire, tandis que Trade Lens est une plateforme de chaîne d'approvisionnement permettant le partage d'informations et la collaboration entre les chaînes d'approvisionnement tout en réduisant les frictions commerciales afin de favoriser le commerce mondial.

L'ensemble des compétiteurs sont limités à des solutions partielles, applicables à un cas de figure précis. A notre connaissance, nous sommes les premiers à offrir une plateforme basée sur la blockchain dans le secteur automobile permettant d'optimiser le processus de gestion des véhicules gravement accidentés.

Concernant notre solution de détection de fragilité financière, notre étude du marché nous a permis de constater qu'il n'existe actuellement aucune solution adressant ce besoin de prédire le taux de résiliation des abonnements pour les opérateurs télécom. Pour leurs contrôles, les opérateurs télécom ont l'habitude d'interroger la base Preventel (résiliation autres opérateurs) et d'analyser la conformité des données bancaires. Les solutions existantes permettent la détection de fraudes financières dans des cas de figures différents au nôtre. Nous pouvons citer à titre d'exemple, la solution **BankWare.NET** qui est un logiciel de détection de fraude financière dédiée aux banques.

Enfin, en ce qui concerne notre solution de prédiction et pilotage des leads, nous avons constaté qu'il existe plusieurs solutions puissantes dans le marché comme les solutions

Commenté [R08]: Pourriez-vous nous citer vos concurrents majeurs ? Comment vous vous distinguez par rapport à leurs solutions ?

TilKee et Salesforce. Tilkee est un logiciel permettant de mieux gérer les propositions commerciales afin d'augmenter les probabilités de conclure une vente tandis que Salesforce est une solution CRM permettant de gérer les leads et les opportunités de ventes. Cependant, ces solutions se limitent à mesurer la contribution de chaque levier sans prise en compte des éléments externes. En contrepartie, notre modèle permet d'intégrer des données externes en prenant en compte l'impact des différentes campagnes commerciales (ex. campagne en ligne vs campagne TV vs campagne radio).

Sur le plan du pilotage, nous nous distinguons par rapport aux solutions existantes par le choix des KPI et l'ergonomie de la solution. Celle-ci a reçu le prix du Data Festival en qualité de meilleur dispositif de pilotage.

1.3 Description des travaux effectués

En 2020, nous avons travaillé sur diverses solutions technologiques afin de proposer des solutions innovantes et adaptées aux besoins de ses clients agissant dans différents secteurs d'activités.

1.4 Blockchain

L'objectif du projet consiste à développer une application pour notre client BCA Expertise, une entreprise d'expertise automobiles qui intervient sur toute la nature des dommages qu'un véhicule a subi lors d'un accident. L'application permet de faciliter le processus de gestion des véhicules gravement accidentés en faisant collaborer les différentes parties prenantes (Garage, Assurance, Experts, Préfecture) grâce à une plateforme technique basée sur la technologie "Blockchain".

Le but est alors de créer un réseau d'acteurs pour diminuer considérablement les frais et les délais d'expertise de transmission de documents, avoir des certifications immuables pour éviter les fraudes, avoir un registre commun et transparent (visible par tous les acteurs de la blockchain) et aussi réduire les coûts et le travail sans valeur ajoutée.

Contrairement à la situation actuelle où les échanges entre les différents acteurs du processus sont réalisés indépendamment et de façon non structurée, l'application permet d'ordonner les actions de chacun des acteurs et de tracer dans une blockchain comme dans un livre de comptes les décisions prises et les actions effectuées à chaque étape du processus.

Ce type de projet nécessite de base une infrastructure solide en matière de sécurité, de robustesse et de scalabilité. Mais au-delà de ces considérations habituelles, ce projet nécessitait de mettre en œuvre une blockchain qui n'impose pas aux différents acteurs une infrastructure technique (datacenter, cloud) particulière. En effet, afin d'obtenir

l'adhésion de tous les acteurs, il faut pouvoir leur laisser la possibilité de contribuer à la blockchain dans l'environnement qu'ils choisissent.

.1.4.1.1 Configuration de la console

Tout d'abord, nous avons configuré la console IBM blockchain via le cloud IBM qui propose une multitude de fonctionnalités liées aux différents outils proposés. Cette console permet d'ajouter les différents acteurs, les différents privilèges ainsi que les possibilités de transactions.

Pour créer notre propre blockchain, nous avons commencé par créer un service Kubernetes. C'est un conteneur d'applications qui permet un déploiement automatisé dans le cloud IBM, ce qui nous permet de déployer un cluster qui fait office de base de données distribuée. Ensuite, nous avons lancé la plateforme pour décider du consortium, soit les participants du réseau (Voir Figure 1).

Dans notre cas, nous avons défini 3 acteurs notamment : le ministère, les experts automobiles et les compagnies d'assurances. Une fois que nous avons décidé des composants comme le nombre d'organisations présentes, le nombre d'acteurs présents dans chacune d'elles, nous avons affecté «l'ordering Node» et les certificats d'autorités. (Voir Figure 2)

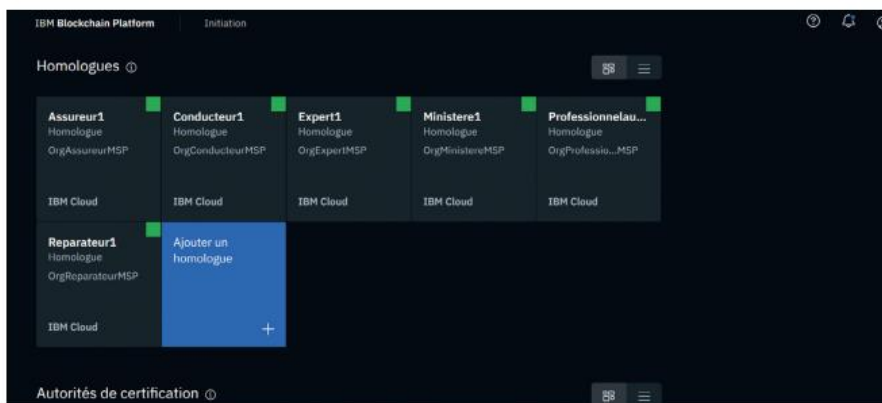


Figure 1 : Différents acteurs de la blockchain

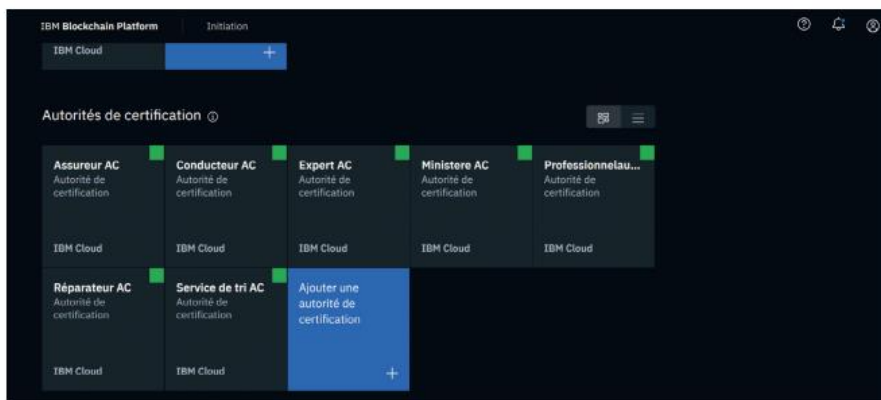


Figure 2 : Affectation des certificats d'autorité

De nombreuses blockchains distribuées n'ont pas ces attributs précédemment cités. C'est-à-dire qu'elles fonctionnent avec des processus de consensus basés sur des algorithmes de probabilités complexes qui garantissent la cohérence des processus par rapport au registre commun à un degré élevé, par exemple Bitcoin ou Ethereum. En revanche l'outil de conception de blockchains privées Hyperledger est composé d'un nœud différent des autres nommé « *ordering Node* ». C'est celui qui gère toutes les transactions et s'assure de leur conformité au sein du réseau. Les certificats d'autorités sont les permissions accordées à certains utilisateurs dans chaque organisation, ce sont les Membership Service Provider (MSP).

Les organisations sont les différentes parties prenantes. Dans le cas de notre PoC, les différentes parties prenantes sont les particuliers, BCA ou les experts automobiles, une entité représentant la préfecture et les assureurs. Nous pouvons enregistrer les différents utilisateurs dans chaque organisation, ce qui nous donnera la possibilité après cette étape de définir les MSP.

.1.4.1.2 Développement du Smart Contract

Les Smart Contract représentent les contrats virtuels ayant principalement un fonctionnement d'évènement déclencheur. Si les conditions d'une fonction sont validées alors il s'exécute.

Dans notre cas, un cas d'usage réalisable serait celui où un expert automobile certifie qu'un véhicule ne peut plus rouler. Alors il aurait la possibilité d'envoyer l'information « le véhicule ne peut plus rouler » qui sera immédiatement transmise chez les assureurs, ce qui annule les contrats d'assurances et facilite la transmission des informations offrant ainsi une réduction des coûts. Effectivement il y a un gain de temps et de ressources entre tous les différents acteurs de la chaîne et les étapes qui les séparent dans la compagnie, de même pour les transmissions de documents au gouvernement. Toutes les étapes avant le smart contract sont facilitées par la console IBM, contrairement aux autres blockchains qui doivent être codées manuellement du

début à la fin. La console permet de générer toutes les étapes énoncées précédemment et de simplifier le processus.


.1.4.1.3 Développement de l'application client

Une fois le smart contract réalisé, nous avons développé l'application « client » permettant de donner un aspect visuel au contrat réalisé en typescript.

Les deux différentes interfaces que nous avons créées sont liées aux différentes fonctionnalités. La première est celle de l'expert automobile, dans celle-ci il pourra remplir différents champs concernant les informations d'un véhicule à expertiser, la date, si la voiture est en état fonctionnel et s'il y a des réparations. La seconde est celle du ministère. Elle permet au gouvernement de certifier les différentes informations transmises par l'expert, pour que l'assureur soit ensuite informé des décisions prises (Voir Figures 3 et 4).

Expert View Ministry View Admin View

Vue expert automobile



Caracteristiques du véhicule

Nom du propriétaire

N° immatriculation

Marque

Couleur

Conclusions de l'expert

Date

Imréparable A réparer Fonctionnel

Non

Oui

Valider

Figure 3 : Vue expert automobile

The screenshot shows a web application interface. At the top, there are three navigation links: 'Expert View', 'Ministry View', and 'Admin View'. Below them is a large blue header bar with the text 'Création d'un véhicule'. Underneath the header, the title 'Ajout d'un véhicule' is displayed. The form contains four input fields: 'Nom du propriétaire', 'N° immatriculation', 'Marque', and 'Couleur'. At the bottom of the form is a green button labeled 'Valider'.

Figure 4 : Vue administrateur

Tout ce processus présente un gain de temps immense entre les différentes organisations ainsi qu'une sûreté sans faille. Effectivement la sécurité des blockchains publiques ou privées permet d'affirmer qu'il est extrêmement difficile de falsifier les informations qu'elles contiennent. La cryptographie est un moyen difficile à inverser, ainsi avec le système d'imbrication entre les différents blocs, cela rend la corruption très complexe.

Avant de déployer l'application, la dernière étape est de relier le smart-contract et la console IBM à l'application client. Cette partie permet de récupérer les informations via les différentes « requêtes » faites depuis les vues clients. Les différentes couches sont dans l'ordre : l'application client, le backend puis la blockchain.

Finalement, nous pouvons voir l'ensemble du projet avec plusieurs entités : l'application client depuis laquelle l'utilisateur peut effectuer des requêtes qui seront envoyées à la deuxième entité, le « backend5 » ou autrement dit, le côté serveur.

Ce côté récupère les différentes requêtes et les traite selon les différentes fonctions développées. Le backend est en communication avec la base de données qui est développée avec MongoDB et avec la partie blockchain et ses smart-contracts. L'architecture globale de la blockchain est illustrée en Figure 5.

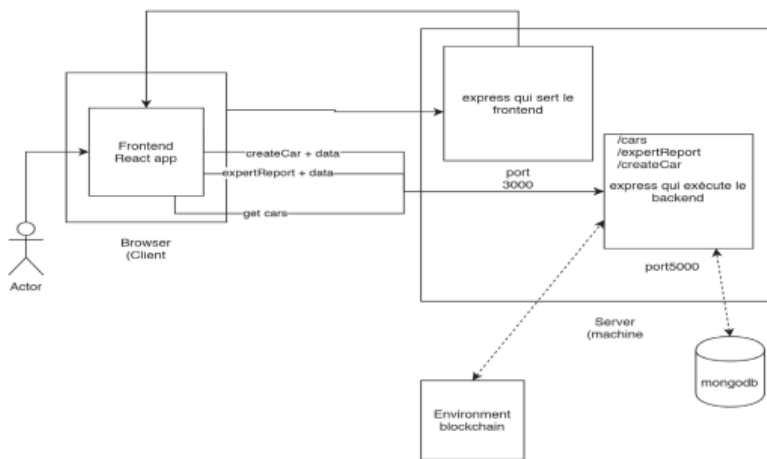


Figure 5 : Architecture globale de la blockchain

Après avoir développé l'intégralité de l'application client et ses différentes fonctionnalités, nous avons déployé l'application sur le « cloud » d'Amazon, cela permet aux consultants de l'entreprise de pouvoir accéder à cette démo sans notre présence et de s'en servir. Il s'agit d'Amazon web services (AWS) qui consiste en un ensemble de services proposés, comme des outils, de la puissance de calcul, des jeux ou encore du stockage. Le cloud est accessible par n'importe qui ayant une bonne connexion Internet.

1.5 Développement d'une architecture Micro-Services IA

L'IES, entité de la société BCA Expertise, opère plusieurs micro-services en interne basés sur l'Intelligence Artificielle dans le domaine de l'assurance automobile notamment, les services de sécurité, supervision, gestion des clients, stockage des données, Backup, service Mesh, développements IA ainsi que le développement des logiciels. Ces micro-services reposent sur des modèles entraînés au sein de l'IES. Ils sont accessibles via une API REST et sont exécutés sous forme de containers docker sur l'infrastructure de BCA Expertise.

Si aujourd'hui les solutions IA de l'IES rendent les services attendus, ils n'offrent pas les meilleures garanties en termes de scalabilité, de gestion multi-clients, de supervision et de bonnes pratiques de déploiement. De plus l'interaction entre les différents services n'est pas conçue de façon optimale. Souhaitant capitaliser sur ces micro-services, l'IES a décidé de les proposer en externe en mode SaaS.

Ainsi, l'objectif de ce projet est de concevoir et mettre en œuvre une nouvelle architecture, capable de supporter la charge induite par ce nouveau positionnement en tant que fournisseur de service et permettant d'intégrer rapidement de nouvelles

offres de services, gérer l'accès et la facturation des clients ainsi que fournir tous les outils nécessaires au suivi de la production, tout en limitant au maximum l'impact de ce changement d'architecture sur les micro-services eux-mêmes.

Pour ce faire, il est nécessaire d'employer un socle technique commun permettant d'opérer des modules IA génériques mais également spécifiques à certains clients. L'unification des processus et l'industrialisation des déploiements permettent d'accepter plus facilement une montée en charge importante et de faciliter le développement de nouveaux services pour des clients de plus en plus nombreux.

Nous avons choisi d'employer le mode de fonctionnement Docker qui permet d'encapsuler un logiciel et toutes ses dépendances (bibliothèque, code source, fichiers de configuration) dans une entité appelée « container ». Ce container peut ensuite être déployé sur n'importe quel serveur disposant d'un environnement d'exécution compatible avec Docker.

Ce mode de fonctionnement facilite grandement le déploiement de logiciels. Avec Docker, le développeur livre un logiciel qui est à la fois « packagé » et prêt à l'emploi. Il ne reste plus à l'équipe d'exploitation de récupérer l'image du container du logiciel, de l'exécuter sur un ou plusieurs serveurs et de gérer les problématiques réseau pour permettre la communication avec d'autres systèmes ou l'accès réseau au logiciel déployé.

La Figure 6 ci-dessous présente l'intégration d'Istio dans un cluster Kubernetes pour implémenter l'architecture micro-services de la plateforme IA de l'IES. Une description détaillée de l'architecture est fournie en Annexe1.

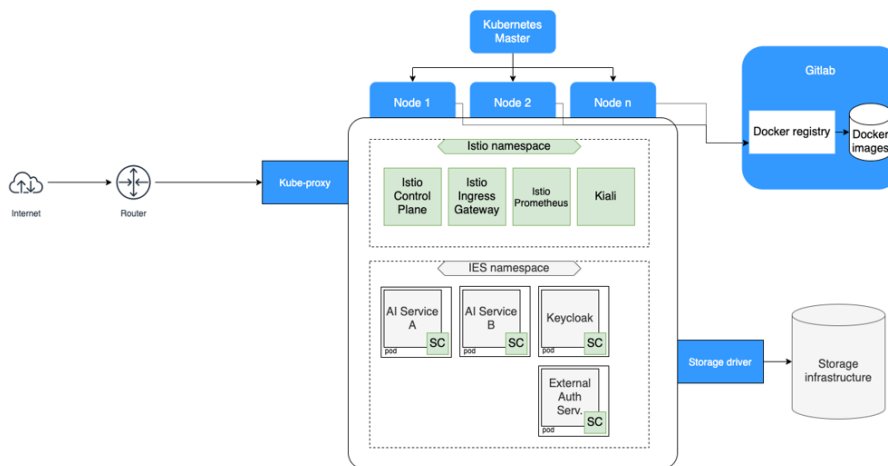


Figure 6: Implémentation de l'architecture micro-services de la plateforme IA de l'IES

L'architecture est conçue à partir de l'association de Kubernetes et Istio.

Kubernetes est une technologie très répandue dans les architectures microservices, en particulier dans les environnements cloud ou pour le déploiement de cloud native applications. Un cluster Kubernetes permet d'industrialiser le déploiement de containers Docker et de répondre à de nombreuses exigences d'un système en production :

- Scalabilité ;
- Optimisation des ressources ;
- Tolérance aux pannes ;
- Standardisation des déploiements (manifestes) ;
- Extensibilité (Custom Resource Definitions) ;
- Stabilité (rolling upgrades).

Istio est un Service Mesh utilisé depuis plusieurs années en production. Outre le fait de jouer le rôle d'API Gateway dans l'architecture microservices, Istio apporte des fonctionnalités qui viennent compléter celles de Kubernetes :

- Gestion et administration centralisée du maillage de services sans impact sur le développement des microservices ;
- Routage dynamique ;
- Contrôle d'accès ;
- Audit / monitoring / métriques / traçage
- Équilibrage de charge entre les multiples instances des microservices ;
- Circuit-breaking en cas d'indisponibilité d'un microservice pour éviter les erreurs à répétition ;
- Politiques de « retry » pour retenter l'appel à un service en cas d'erreur ou d'indisponibilité de celui-ci avant de considérer la requête en erreur.

1.6 Fragilité financière

Dans un contexte de crise sanitaire ayant déjà fragilisé la santé économique de beaucoup d'entreprises, il était important que les équipes de management du risque et du recouvrement redoublent d'efforts afin que leurs clients paient en temps et en heure et n'entraînent ainsi pas de problématiques de trésorerie supplémentaires. Bien prédire, ou mieux prédire, ses futurs mauvais payeurs devient par conséquent un enjeu incontournable.

L'opérateur Télécom La poste Mobile pour lequel l'algorithme a été réalisé commercialise des abonnements avec smartphone. Une part importante de ses clients est en fragilité financière et nécessite d'être relancée tous les mois en vue du paiement des factures. L'opérateur souhaitait mieux connaître le risque client au moment de la souscription afin de demander des garanties ou d'orienter les clients les plus à risque vers des offres plus adaptées.

Le projet vise à limiter les demandes de garanties aux seuls clients les plus à risque ainsi qu'à réduire le taux de résiliation des abonnements à 3 mois et atteindre un taux de 2% (le taux de résiliation actuel étant de 4%). Les gisements de données constituent désormais une opportunité permettant à la fois une connaissance approfondie des clients et prospects, mais en même temps une difficulté accrue pour les équipes d'analyser les risques et de proposer une méthode harmonisée de prévention et de gestion des impayés.

Les actions possibles sont nombreuses et complexes à coordonner pour être ajustées à chaque niveau de risque client : orientation sur des moyens de paiement moins risqués, recours à des solutions de scoring de fraude selon le canal de vente, optimisation de la date de prélèvement, pratique d'une avance sur consommation avant la contractualisation, limitations provisoires de service, rééchelonnement des montants dus, accompagnement par le conseiller clientèle vers des niveaux de services plus adaptés à la situation de l'abonné.

Afin de mener à bien cette mission, l'implémentation d'algorithmes de prédiction du niveau de risque pour un client à générer un impayé passe par les étapes suivantes.

.1.6.1.1 Définition de la problématique et de l'indicateur métier

Contrairement à l'acte de fraude qui est parfois ambigu et difficile à déceler, l'impayé est un acte immuable. La qualité d'un client, pour lequel un historique à être « bon payeur » ou « mauvais payeur » est disponible, peut donc être évaluée assez rapidement. Ainsi, une approche supervisée est privilégiée plutôt qu'une méthode non supervisée souvent moins performante.

Afin de délimiter complètement la base de données d'apprentissage, il faut déterminer l'historique nécessaire de données étiquetées. Cette fenêtre de temps nécessaire dépend du volume de nouveaux abonnés mensuels de l'entreprise, et est fonction de la capacité à mettre à disposition ces données étiquetées. Il va de soi que l'apprentissage du modèle prédictif est d'autant meilleur que la base de cas de « bon payeur » et « mauvais payeur » est étayée. Il est également intéressant d'avoir un historique suffisamment long pouvant faire apparaître des effets de cycles ou saisonnalité. Pour ces raisons, nous avons donc choisi de nous baser sur un historique de deux ans de données clients pour notre approche de scoring avec l'opérateur téléphonique.

.1.6.1.2 Pré-traitement et sélection des données pertinentes

Prédire le caractère « mauvais payeur » suppose d'avoir les données les plus pertinentes possibles. Voici quelques exemples de types de données qu'il est intéressant d'intégrer selon le cas pour avoir une vision complète des abonnés :

- Données sociodémographiques : âge, sexe, nationalité, langue, situation civile et familiale, CSP, lieu de vie, etc. ;

- Données comportementales liées aux goûts, choix et habitudes de consommation : historique des produits achetés ou services souscrits, options souscrites, fréquence d'achat et niveau d'utilisation des services, canal de souscription, etc. ;
- Données financières : banque, choix du mode de paiement, jour de prélèvement, etc. ;
- Données externes et enrichissement via des *open datas* :
 - o Données macro-économiques territoriales :
L'INSEE publie tous les trois ou quatre ans un certain nombre d'indicateurs (éducation, chômage, revenus, démographie) à l'échelle de l'IRIS par exemple. Les *Îlots Regroupés pour l'Information Statistique* sont une maille élémentaire infra-communale de découpage du territoire français qui correspondent à des sortes de quartiers d'environ 2 000 habitants.
 - o Données épidémiologiques:
La conjoncture sanitaire actuelle nous force à changer notre paradigme et notre façon de penser afin d'intégrer le risque de crise sanitaire dans la prise de décision. La souscription d'un service ou sa mise en œuvre peut être fortement impactée selon le département. Il est aussi nécessaire de prendre en compte leur impact pour actualiser le modèle, notamment celui sur les effets liés aux séries temporelles.

Une fois que toutes les variables jugées pertinentes par la prédiction ont été définies, nous utilisons des méthodes de réduction de dimensions ou de sélection de variables afin de ne sélectionner que les variables qui sont réellement les plus pertinentes statistiquement parlant.

Les méthodes de réduction dimensionnelle de type PCA (*Principal Component Analysis* ou ACP pour *Analyse en Composantes Principales* en français) sont très pratiques mais créent des combinaisons linéaires des variables existantes qu'il est parfois difficile d'interpréter. C'est pourquoi, pour une meilleure interprétabilité des résultats, nous avons privilégié l'utilisation d'une méthode de sélection de variables : tests statistiques (ANOVA, Chi2, t-test de Student...), suppression de variables colinéaires, *Forward selection*, *Backward elimination*, etc.

.1.6.1.3 Apprentissage automatique

Lors de la phase du scoring du risque d'être « mauvais payeur », nous avons opté pour des méthodes ensemblistes et testé des algorithmes de *Random Forest* et *XGBoost*. C'est finalement le *Random Forest* qui nous a donné les meilleurs résultats.

Commenté [R09]: Pourriez-vous nous communiquer les résultats obtenus par l'algorithme Random Forest ?

1.7 Prédiction et pilotage des leads

L'objectif du projet est de développer une application pour notre Client SKODA permettant de prédire le nombre de leads qualifiés dans le secteur de l'automobile en fonction du plan marketing (publicité, messages, budget par partenaire), des éléments

externes (ex. nombre de cas de COVID) ainsi que des caractéristiques de la gamme (modèles, âge, etc.).

L'originalité du projet réside dans l'intégration de données externes, la prise en compte du plan marketing ainsi que la sélection des leads qualifiés à la place des leads d'une faible probabilité de transformation.

Sur le plan du pilotage, le caractère innovant réside dans le choix des KPI et l'ergonomie de la solution. Celle-ci a reçu le prix du Data Festival comme meilleur dispositif de pilotage.

Pour se faire nous avons commencé par la collection et la préparation des données. Tout d'abord, nous avons déterminé une date de début et de fin pour les campagnes fil rouge performance par la récupération des données sur les impressions. Nous avons récupéré par la suite des budgets par campagne fil rouge performance. Nous avons toujours des discussions en cours avec Remind sur les délais de mise à disposition de ces données.

Concernant l'algorithme d'apprentissage, nous avons choisi les modèles ensemblistes Random Forest et XGBoost vu leur bonne « explicabilité » et la précision des prédictions que ces algorithmes produit, associés au fait qu'ils permettent d'éviter le surapprentissage.

Afin d'évaluer nos modèles, nous avons comparé les résultats des algorithmes avec une valeur de référence (baseline) qui correspond à une prédiction évidente. Nous obtenons un coefficient R2 de 0,8, rendant les deux modèles satisfaisants. Le tableau ci-après détaille les résultats obtenus.

Modèle	MSE = Mean Squared Error (écart-type)	MAE=Mean Absolute Error (écart-type)	R2=coefficient de corrélation (écart-type)
Baseline (M=M+1)	7418	47	0.60
Random Forest	3643 (865)	32 (4)	0.79 (0.08)
XGBoost	3564 (643) -52%	32 (4) -32%	0.79 (0.08) +32%

Figure 5 : Résultats des algorithmes Random Forest et XGBoost

Baseline choisi : Le volume de leads qualifiés par partenaire et levier prédit correspond au volume de leads du mois M-1 pour le même levier et partenaire.
Moyenne mensuelle des leads : 100

1.8 Conclusions

Soucieux de la qualité de nos solutions, nous restons à l'écoute de nos clients et faisons preuve de proactivité pour anticiper leurs besoins afin de proposer des expériences ergonomiques et des performances techniques toujours améliorées.

Les travaux menés en 2020 nous ont permis de développer quatre outils technologiques spécialisés pour nos clients dans les domaines de télécom et assurance automobile.

Des considérations d'exhaustivité des fonctionnalités, d'ergonomie d'usage et de performance ont guidé nos développements. Le choix de nouvelles technologies (langage de programmation, hébergement...) sert également notre objectif de différenciation par rapport à nos concurrents grâce à des plateformes hautement automatisées et robustes, capables d'offrir les meilleures performances tout en évoluant avec un volume de demandes en croissance.

ANNEXES

Annexe 1 : Plateforme micro-services IA

Annexes

Annexe 1 : description de Predict Value

Annexe 2 : Guide technique pour l'amélioration de Predict Value