IA et procédés industriels : vers des modèles explicables en soutien à l'expertise humaine

Baptiste Fedi, fondateur et PDG d'Hivelix

L'intelligence artificielle est plus que jamais sur toutes les lèvres, notamment grâce à l'accessibilité des grands modèles de langage (LMM) comme Chat GPT d'Open AI, pour citer la plus connue, ou les modèles de Mistral AI, start-up française souhaitant faire sa place face aux GAFAM. Cet engouement autour des capacités de ces modèles génère ou ranime par la même occasion de nombreux fantasmes en termes de capacités, mais également de craintes, toutes deux en partie infondées comme cela a pu être le cas face aux balbutiements des premiers concepts d'intelligence artificielle ou de nombreuses technologies de rupture.

Face aux fantasmes autour de ces modèles, il est bon de rappeler que de nombreux experts de l'IA s'accordent à dire que ces modèles peuvent être pertinents et trouveront leurs usages, mais que les algorithmes sous-jacents ne sont en rien similaires au cheminement intellectuel humain, limitant les réponses pertinentes dans le cadre d'actions nécessitant adaptabilité et créativité, indispensables à l'innovation.
Luc Julia[1], expert français de l'IA et auteur du livre au titre volontairement provocateur mais explicite : « L'intelligence artificielle n'existe pas », préfère utiliser le terme « augmentée » à la place d'artificielle. De façon annexe, Yann LE CUN, ayant reçu le prix Turing en 2018 pour ces travaux sur l'IA, considère que « l'IA amplifie l'intelligence humaine comme la machine amplifie la force »[2].
Cet usage du terme « Intelligence Augmentée » traduit le fait que l'ensemble des algorithmes aujourd'hui à disposition (apprentissage automatique supervisé, non supervisé, apprentissage profond, LMM…) ne sont que des outils numériques qu'il est bon d'utiliser à bon escient avec l'approche adaptée afin d'améliorer nos capacités cognitives d'analyses ou de capitalisation de données, comme nous le faisons déjà naturellement avec la simulation numérique. Dans le domaine de l'ingénierie des matériaux ou des procédés industriels, comme nous allons l'évoquer ici, ces méthodes doivent faire partie d'un processus complet. Ce processus inclut la génération de base de données (caractérisation, suivi de production…), le monitoring de ces dernières, le traitement par des méthodes numériques en complémentarité et non en opposition avec l'expertise humaine. En effet, bien que ces algorithmes ne soient pas intelligents à proprement parler, ils ont en revanche des capacités d'analyses de grands volumes de données et d'identification de schémas de corrélation complexes entre des paramètres dont il est bon de tirer parti.

Cependant, les modèles d'IA, avec leur opacité, peuvent donner le sentiment d'être des « boîtes noires » vis-à-vis de leur fonctionnement et de la manière dont les données et leur interdépendance ont été assimilées par l'algorithme quel qu'il soit. Dans le cas d'une régression linéaire, la compréhension est simple, mais dans le cas d'arbres de décision ou d'un réseau de neurones complexe, les mécanismes de décisions successifs ne sont pas triviaux. Face à cela, nous verrons comment l'explicabilité des modèles de Machine Learning (ML) peut être utilisée afin d'appréhender la manière dont l'algorithme a interprété les données. L'explicabilité des modèles est un élément précieux à ne pas négliger afin de faciliter l'intégration de ces outils. En effet, la compréhension de l'approche numérique peut ensuite être comparée à un avis d'expert et à sa compréhension intuitive. Nous allons voir comment ces approches peuvent être appliquées dans le cas d'un procédé de traitement de surface, mais également à tout autre procédé industriel.

I. Récoltes et visualisation des données

1. Capitalisation de données

La réalisation d'un dépôt électrolytique permettant d'aboutir à un revêtement conforme en termes d'épaisseur, d'aspect ou encore de composition dans le cas d'un alliage, résulte d'une optimisation multi-paramètres. Ces paramètres sont liés à la formulation composée de plusieurs éléments ainsi qu'aux paramètres procédés appliqués comme la densité de courant, la température ou encore le temps de traitement. Cette optimisation multi-paramètres est classiquement réalisée via un processus d'essais-erreur, traduit sous le terme de savoir-faire, permettant d'ajuster la chimie ou les paramètres procédés. Les approches d'exploitation de données et Machine Learning peuvent être utiles afin de converger plus rapidement vers la combinaison de paramètres optimale ou accélérer la phase d'ajustement d'un procédé hors gamme.

Dans le cas d'un électrolyte de Zinc-Nickel alcalin, la formulation peut être décrite par les concentrations en nickel et en zinc (définies par un ratio Zn/Ni en solution) ainsi que les concentrations en soude NaOH, en différents additifs (appelés Add1 et Add 2) et en composés polluants (pollutions et carbonates). La combinaison des différents paramètres de formulation et des paramètres procédés associés permet d'aboutir à une épaisseur et une composition spécifique, sur une gamme plus ou moins large de paramètres en fonction de la robustesse d'électrolyte.
Dans un environnement industriel, des données sont générées en permanence par différents services (laboratoire d'analyses, production, qualité…) et peuvent être croisées et exploitées afin d'en extraire de la valeur. Un exemple théorique de données générées par un atelier de

traitement de surface sur une période de trois ans est détaillé sur les **figures 1 et 2**. Ces données sont, en général, dans une gamme définie par une fiche technique (**tableau 1**), comme étant une gamme suffisamment robuste afin d'atteindre les spécifications attendues pour le revêtement (**figure 1 Atelier**). Cependant, il peut arriver qu'à la suite d'une dérive du bain ou d'une erreur humaine, les paramètres (chimie, paramètres procédé…) soient en dehors des gammes (**figure 1 Labo**). La Figure 1 décrit les effectifs en nombre de valeurs de chacun des paramètres étudiés ici.

La **figure 2** présente les épaisseurs et teneurs en nickel des revêtements en fonction des données d'entrées détaillées dans la **figure 1**.

2. Visualisation de données

À la suite de la récolte des données, une phase de visualisation de ces dernières est indispensable. Cette étape permet de transformer un ensemble de données complexe en informations compréhensibles afin d'identifier des tendances ou des motifs, mais également des anomalies. Cette étape permet d'identifier également des lacunes ou biais dans les données récoltées ou encore d'identifier des méthodes de pré-traitement adéquates avant de passer à une phase d'apprentissage via des algorithmes de Machine Learning.

i. Diagramme de coordonnées parallèles

La représentation sous forme d'un diagramme de coordonnées parallèles, comme représenté sur la **figure 3**, est une visualisation intéressante permettant de comparer plusieurs variables de manière simultanée. Cette représentation permet de visualiser des relations complexes, des clusters de données ou de comportement ou encore d'identifier des points aberrants rapidement.

Ces diagrammes sont des représentations interactives permettant de naviguer aisément au sein des données afin d'analyser les liens de causalité entre données d'entrée et paramètres de sortie.

Dans cet exemple, il apparaît que les dépôts les moins épais sont ceux possédant la teneur en nickel la plus élevée. Ce paramètre est également lié à différents paramètres d'entrée. En effet, les concentrations de chacun des constituants impactent logiquement les données de sortie tout comme la température et la densité de courant appliquée.

Composants	Optimum	Minimum	Maximum
Ratio Zn/Ni	**6,67**	5,4	8,2
[NaOH] (g/L)	**130**	100	150
[Add 1] (g/L)	**60**	50	80
[Add 2] (g/L)	**4,8**	3,6	8,4
Température (°C)	**25**	20	26
Densité de courant (A/dm²)	**5**	3	6
Pollution (µg/L)	**0**	0	100
Carbonates (g/L)	**0**	0	60

Tableau 1 : Valeurs optimales et gammes de chacun des paramètres d'un électrolyte de ZnNi alcalin

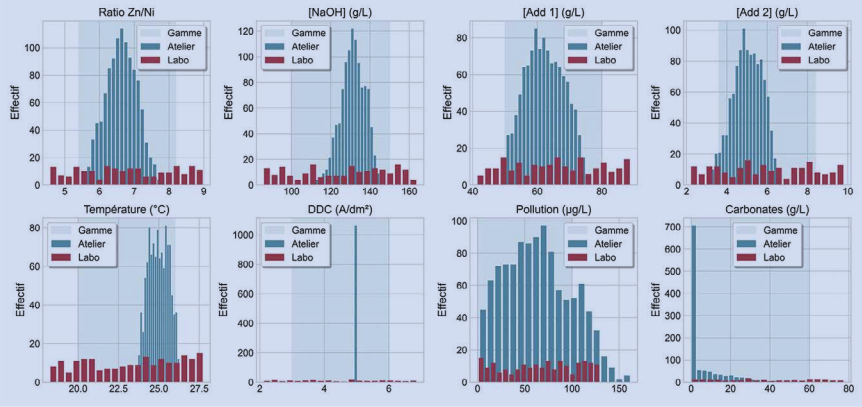


Figure 1 : Histogrammes de répartition des effectifs des valeurs générées pour chaque paramètre d'entrée.

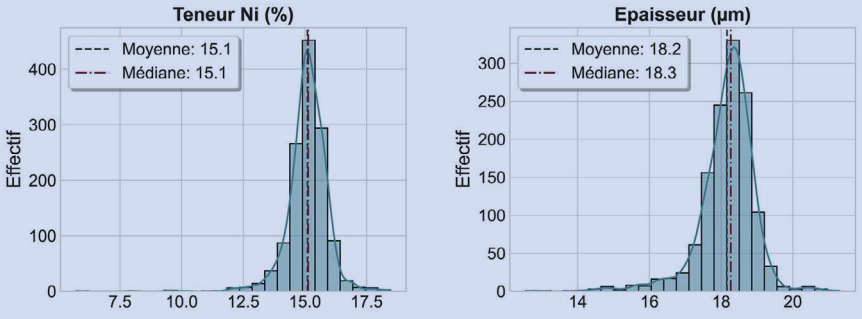


Figure 2 : Histogrammes de répartition des effectifs des données de sortie calculées.

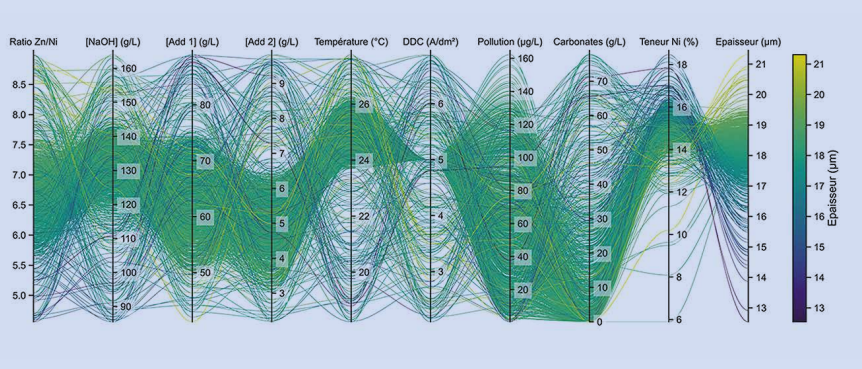


Figure 3 : Diagramme de coordonnées parallèles

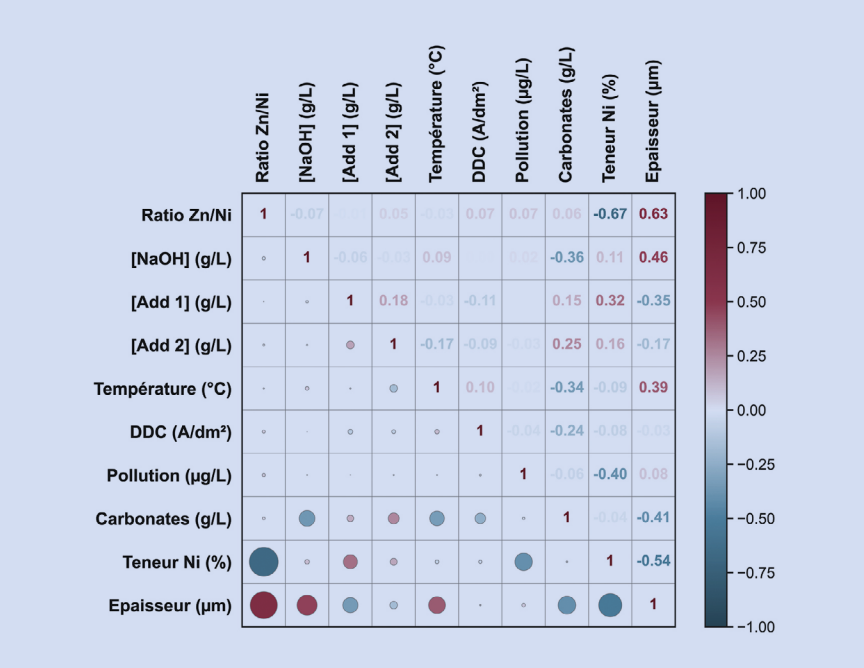


Figure 4 : Heatmap de corrélations (méthode de Pearson)

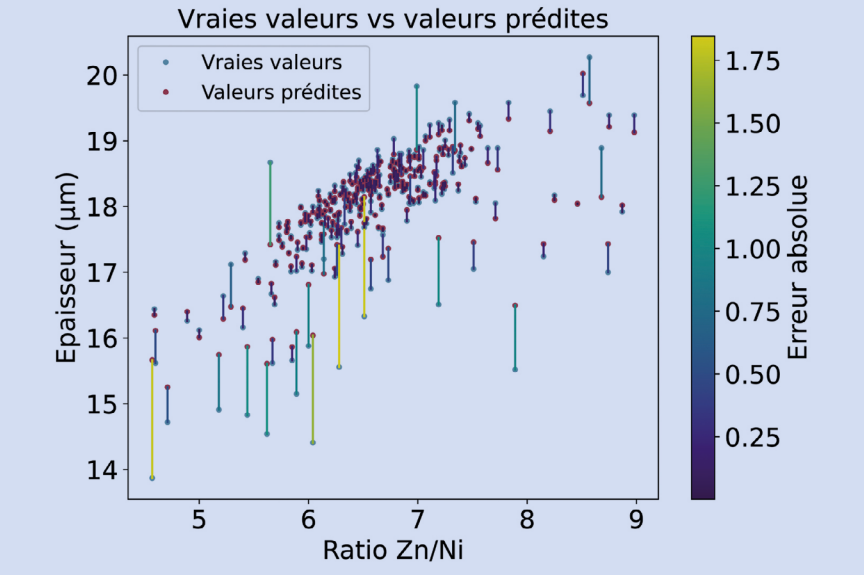


Figure 5 : Comparaison des valeurs prédites avec les valeurs réelles de l'ensemble de test

Pour générer un modèle de prédiction, différents types de modèle de ML peuvent être utilisés. Ces modèles sont répartis en différentes catégories d'apprentissage : supervisé, non supervisé, semi-supervisé et par renforcement.

ii. Heatmap

Une autre visualisation graphique pertinente est la représentation sous forme de heatmap. Cette représentation permet de quantifier les corrélations via différentes méthodes (Pearson, Spearman ou encore Kendall). La **figure 4** présente un exemple créé à partir de la méthode de Pearson dont les coefficients de corrélation correspondent à une mesure statistique de la relation linéaire entre deux variables. Cette visualisation fait apparaître chaque paramètre de manière opposée aux autres paramètres afin de corréler et quantifier son impact sur chacun d'entre eux. Les coefficients de corrélation varient entre – 1 et 1 suivant qu'il s'agit de corrélations linéaires positives ou négatives. Une valeur proche de 0 indique que la corrélation entre les paramètres est faible ou alors, avec cette méthode, qu'elle n'est pas linéaire. Ceci permet de visualiser rapidement l'intensité et la direction des relations entre les variables.

II. Machines Learning / optimisation de modèle de ML

Une fois les données récoltées, visualisées et nettoyées, la phase de création d'exploitation via des algorithmes de Machine Learning peut commencer. L'objectif de cette démarche est d'aboutir à un outil d'aide à la compréhension et à la maîtrise du procédé à partir de l'historique des données générées au cours des mois et années d'un procédé industriel. À partir des données d'entrée et de sortie disponibles, il est possible de réaliser un apprentissage automatique pour aboutir à un modèle capable de prédire les données de sortie à partir d'un nouvel ensemble de données d'entrée. Dans notre cas, le modèle devra être capable d'appréhender l'impact de la formulation et des paramètres procédés sur les propriétés du revêtement obtenu (épaisseur, composition).

Pour générer un modèle de prédiction, différents types de modèle de ML peuvent être utilisés. Ces modèles sont répartis en différentes catégories d'apprentissage : supervisé, non supervisé, semi-supervisé et par renforcement. Ici, seul l'apprentissage supervisé est étudié. Le terme « supervisé » est employé, car le modèle apprend les relations entre les paramètres en se fondant sur les sorties de l'ensemble de données.

Pour ce type de données, nous allons chercher à entraîner un modèle de régression. Plusieurs types de modèles sont testés pour déterminer le modèle le plus

performant. Les modèles sont évalués en fonction de leur score de précision obtenu sur l'ensemble des données de test.

Afin d'entraîner un modèle de ML et s'assurer de sa bonne performance, l'ensemble des données est souvent réparti en deux sous-ensembles. Un sous-ensemble contenant 80 % des données destiné à entraîner le modèle et les 20 % restants pour permettre de tester et évaluer la performance du modèle. Cette méthode est souvent efficace, mais dans le cas où l'ensemble de tests contient des informations non couvertes par l'ensemble d'entraînement, le modèle ne peut pas apprendre à gérer ces cas précis et peut résulter en un score qui ne reflète pas réellement la performance du modèle. Pour cela, la solution consiste à utiliser un processus de validation croisée destiné à entraîner et tester le modèle à plusieurs reprises sur des sous-ensembles d'entraînement et de test qui varient.

L'épaisseur et la teneur en nickel sont deux variables continues que nous cherchons à prédire. Il s'agit donc d'un problème à sorties multiples où les variables de sorties sont interdépendantes. Tous les modèles de ML ne sont pas compatibles avec des problèmes à sorties multiples, mais pour ceux-là, il reste possible de considérer chaque variable de sortie comme étant indépendante des autres et ainsi entraîner un modèle à sortie unique sur chaque variable de sortie.

Plusieurs modèles de ML ont été entraînés sans chercher à optimiser leurs paramètres dans un premier temps. Le modèle qui affiche la meilleure précision est un modèle « Extra Trees » faisant partie de la famille des arbres de décision. L'entraînement de ce modèle sur l'ensemble de données générées résulte en un score R^2 d'environ 90 %. La **figure 5** montre ce que représente un score R^2 de 90 % sur l'erreur relative maximale entre les valeurs réelles d'épaisseur de l'ensemble de test et les valeurs prédites par le modèle. Un score de 90 % montre donc une précision relativement intéressante surtout dans les zones où la densité de points est élevée.

L'erreur absolue est très faible dans les zones où il y a beaucoup de données, cependant pour certaines prédictions, l'erreur absolue atteint près de 2 µm. Pour cette raison, il est nécessaire d'améliorer la précision du modèle. Il est possible de réaliser des opérations de préparation de l'ensemble de données afin d'augmenter la performance du modèle. La recherche des paramètres opti-

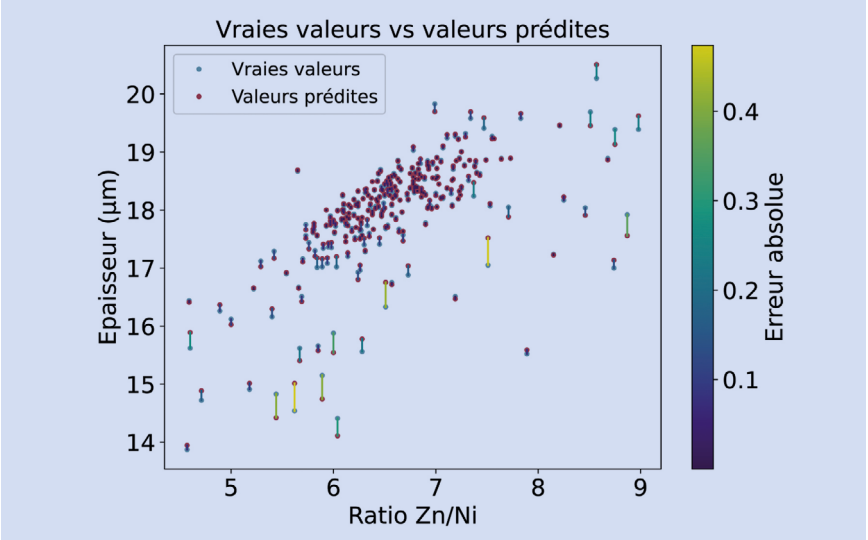


Figure 6 : Comparaison des valeurs prédites avec les valeurs réelles de l'ensemble de test après amélioration du modèle

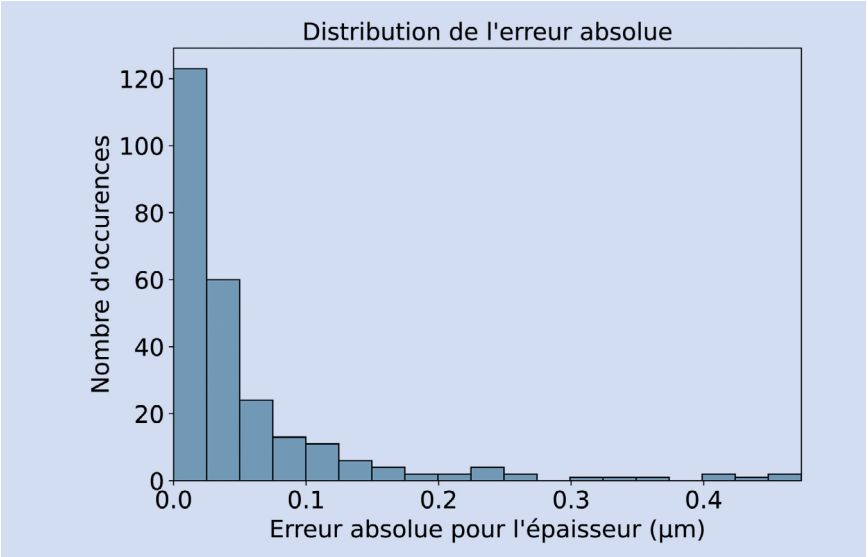


Figure 7 : Distribution de l'erreur absolue sur l'épaisseur

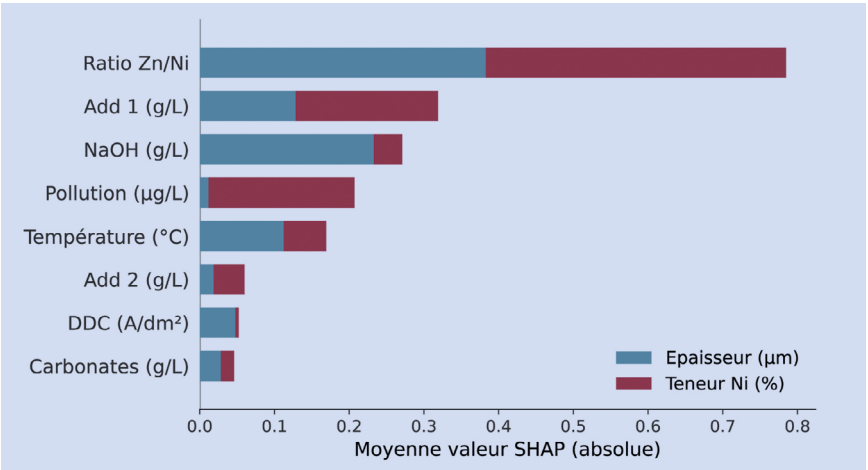


Figure 8 : Impact des données d'entrée sur les données de sortie

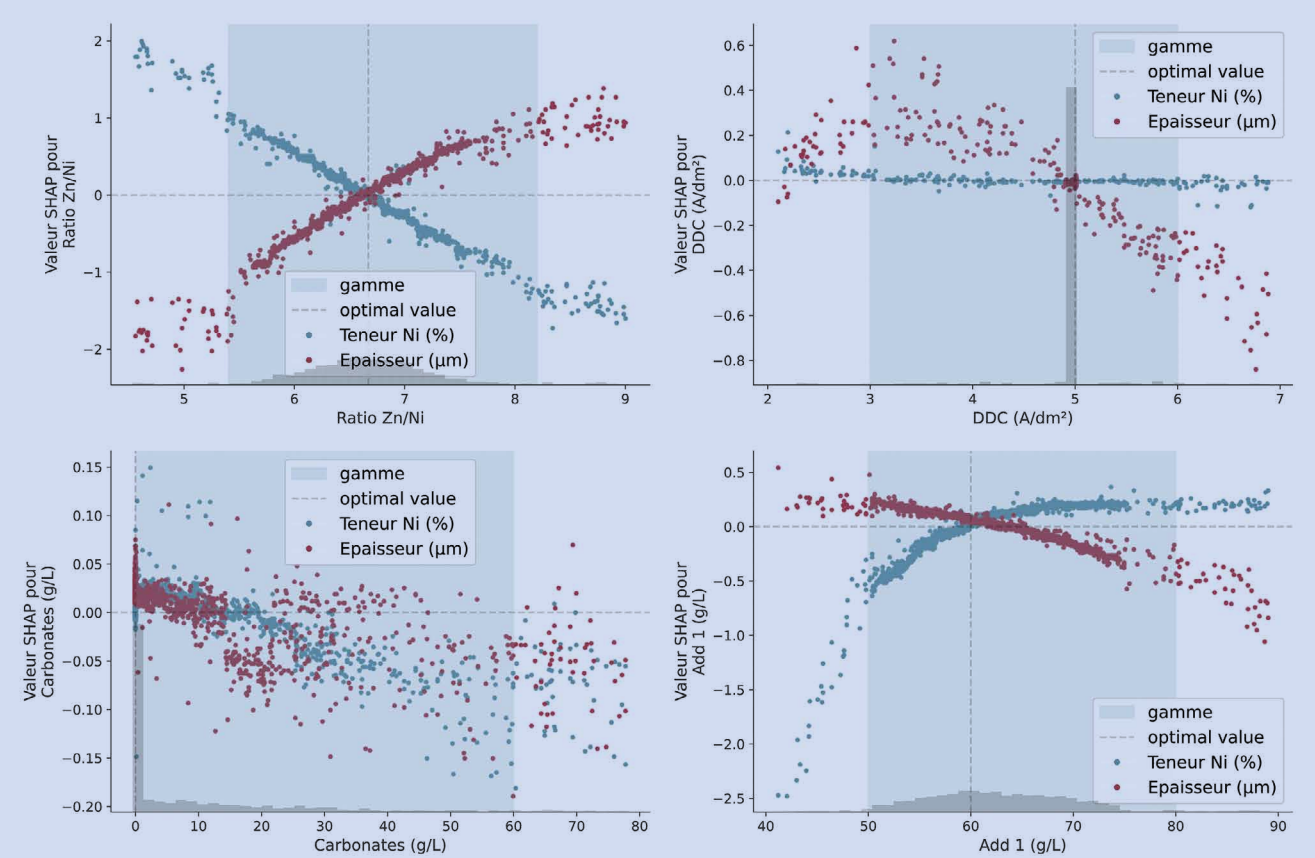


Figure 9 : Valeurs SHAP de chaque paramètre d'entrée pour la teneur en nickel (en bleu) et l'épaisseur (en rouge)

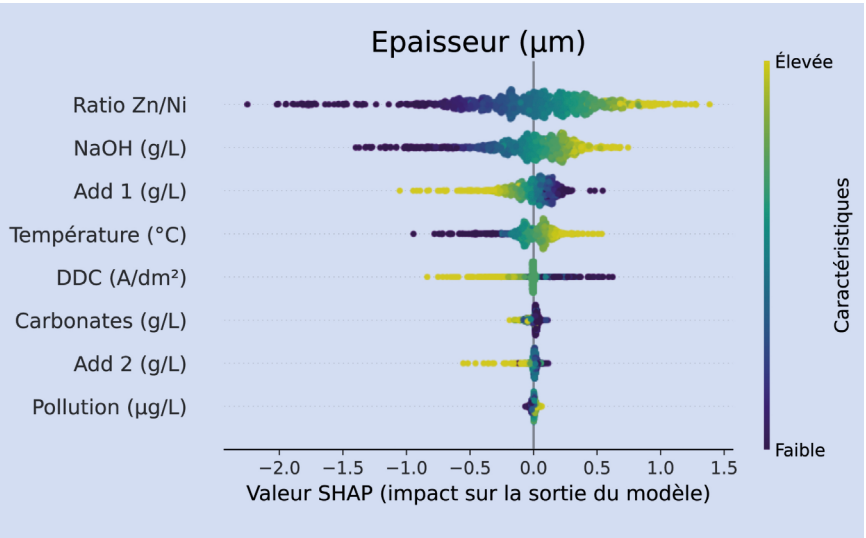


Figure 10 : Valeurs SHAP pour chaque paramètre d'entrée sur la teneur en nickel et l'épaisseur prédites

DDC (A/dm²)	Ratio Zn/Ni	[Add 1] (g/L)	[Carbonates] (g/L)	Température (°C)	[NaOH] (g/L)	[Add 2] (g/L)	[Pollution] (µg/L)
4,07	6,83	69,1	0	24,27	130,25	6,04	2,72

Tableau 2 : Jeu de données utilisé pour la prédiction d'épaisseur détaillée sur la Figure 11

maux (hyperparamètres) du modèle est une étape indispensable afin d'améliorer la précision du modèle. De plus, dans ce cas, il apparaît qu'il est plus efficace d'entraîner un modèle pour prédire la teneur en nickel et un autre modèle pour prédire l'épaisseur tandis que dans le cas précédent, il s'agit d'un modèle unique prédisant les deux résultats à la fois. Après avoir procédé à ces étapes, la précision est nettement améliorée grâce à un modèle de « Gradient Boosting » (**figure 6**). L'erreur absolue maximale est réduite à 0,5 µm et l'erreur moyenne montre que le modèle est très précis avec seulement quelques points hors des données standards qui provoquent une erreur plus importante liée à une quantité de données d'apprentissage réduite dans cette gamme de paramètres. La **figure 7** montre la distribution de l'erreur qui est, dans la plupart des cas, très faible.

III. Explicabilité des modèles d'IA

Les modèles de ML détaillés dans le paragraphe précédent permettent de prédire les propriétés d'un revêtement en se fon-

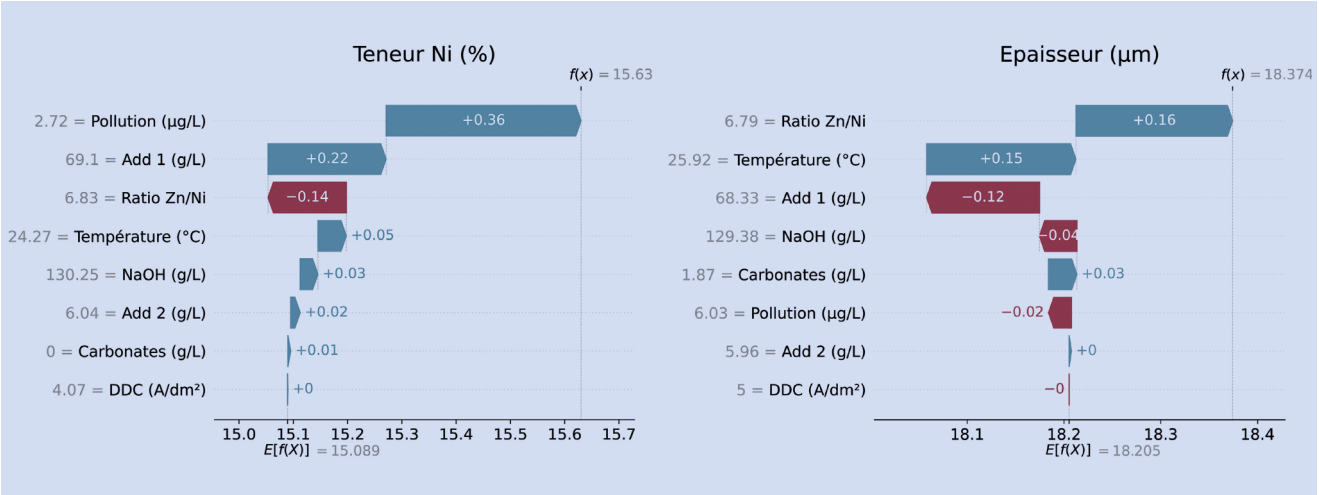


Figure 11 : Explication de la prédiction de teneur en nickel et d'épaisseur pour un seul cas

dant sur les paramètres d'entrée liés à la chimie de l'électrolyte et aux paramètres procédé. Cependant, la manière dont les données sont intégrées et traitées par l'algorithme reste souvent obscure. Il est donc important de pouvoir accéder aux explications sur les chemins de décision de l'algorithme. Cette transparence est essentielle, non seulement pour confronter ces décisions à l'expertise humaine, mais également pour mieux comprendre les éventuelles défaillances et faciliter son optimisation.

Dans le cas d'un simple modèle de régression linéaire, les coefficients de chaque paramètre d'entrée indiquent leur influence sur le résultat de sortie. En revanche, pour des modèles plus complexes, il devient vite difficile de comprendre comment le modèle prédit les résultats, entraînant un effet « boîte noire ». Pour les modèles de type Random Forest basés sur des ensembles d'arbres de décision, il est possible d'utiliser des méthodes d'explicabilité de l'intelligence artificielle (XAI : eXplainable Artificial Intelligence) comme celle des valeurs SHAP (SHapley Additive exPlanations).

Les valeurs SHAP représentent la contribution de chaque caractéristique à la prédiction finale du modèle utilisé. Elles permettent de mieux comprendre comment le modèle utilise les données d'entrée pour prendre des décisions offrant ainsi une transparence et une interprétabilité du modèle essentielles pour maîtriser le procédé. Ces valeurs sont calculées en s'appuyant sur un concept de la théorie des jeux où l'objectif est d'attribuer une contribution à chaque joueur sur le résultat final. Toutes les combinaisons de



Figure 12 : Visualisation des données via l'application



Figure 13 : Prédiction du modèle et explication des prédictions

> Les modèles de ML et l'explicabilité associée peuvent être d'une grande aide à la maîtrise des procédés industriels.

sous-ensembles de paramètres sont étudiées pour calculer la contribution marginale de chaque paramètre absent du sous-ensemble. Par exemple, à partir des valeurs SHAP, il est possible de montrer l'importance de chaque paramètre d'entrée sur les données de sortie du modèle (**figure 8**).

Ainsi, il apparaît que le ratio Zn/Ni est le paramètre qui semble jouer le rôle le plus important à la fois sur la valeur prédite de teneur en nickel et sur celle de l'épaisseur. La concentration de soude (NaOH) a un impact important sur l'épaisseur, mais pas sur le teneur en nickel alors que pour la pollution, le phénomène inverse est observé.

La **figure 9** représente les valeurs SHAP en fonction de quatre paramètres sur la teneur en nickel (en bleu) et l'épaisseur (en rouge). Ce graphique met en évidence un effet inverse du ratio Zn/Ni sur la teneur en nickel et l'épaisseur. La concentration en additif 1 a un effet négatif sur la teneur en nickel lorsque la concentration est en deçà de la valeur minimale de la gamme (50 g/L). Dans la gamme, ce paramètre a un effet relativement faible (valeur SHAP proche de zéro). Plus généralement, les tendances définies lors de la génération des données apparaissent grâce à l'analyse des valeurs SHAP.

La figure 10 montre la distribution des valeurs SHAP pour chaque paramètre d'entrée. La couleur indique la valeur du paramètre (valeurs élevées en jaune, valeurs faibles en violet). Par exemple, une valeur élevée du ratio Zn/Ni a une influence négative importante sur la teneur en nickel, alors que pour la concentration en additif 2, une valeur élevée aura un impact positif. Les données d'entrée apparaissent par ordre d'importance. Ainsi, cet ordre est différent pour les prédictions de teneur en nickel et pour l'épaisseur.

Si l'on prend un peu de recul sur ce type de représentation, nous constatons que cela permet de visualiser de manière explicite comment l'algorithme de ML a interprété le nuage de point qui lui a été fourni, afin de la comparer à une interprétation intuitive humaine. Les corrélations indiquées sont en accord avec le savoir-faire du métier en termes d'impact de la formulation ou des paramètres procédés.

Les figures précédentes fournissent des indications pour expliquer de manière générale comment fonctionne le modèle entraîné. Il est aussi possible d'expliquer localement les prédictions du modèle. C'est-à-dire que pour une instance donnée (celle du **tableau 2** par exemple), les valeurs SHAP permettent d'expliquer comment ont été prédites les données de sortie. La **figure 11** montre l'influence directe de chaque paramètre d'entrée sur les prédictions. Pour lire la figure, il faut commencer par la valeur E[f(X)] qui correspond à la valeur moyenne de l'épaisseur sur l'ensemble de données. Puis, petit à petit, chaque paramètre d'entrée va apporter sa contribution positive ou négative pour arriver finalement à une épaisseur prédite de 18,591 µm.

IV. Interface IA-Humain : Logiciel d'aide à la décision

Ce cas d'exemple démontre la capacité d'un algorithme à capter les dépendances des paramètres sur les propriétés en sortie, celui-ci peut être utilisé comme un outil d'aide à la décision. Il peut également faciliter le suivi et l'ajustement du procédé, sa compréhension, et être une aide précieuse à la résolution de problème ou de crise via l'encapsulation de l'historique du procédé.

Afin de faciliter son usage comme aide à la décision et faciliter l'analyse de grandes quantités de données, un logiciel dédié peut être développé afin d'explorer l'historique des données et prédire les propriétés cibles. L'interface permet à l'utilisateur de visualiser et naviguer à travers la base de données pour faciliter l'analyse des données ainsi qu'utiliser le modèle entraîné pour prédire les données de sortie en fonction des données d'entrée. L'interface graphique permet également de visualiser les résultats de manière intuitive et d'expliquer les prédictions du modèle obtenus grâce aux méthodes d'explicabilité des modèles. L'onglet visible sur la **figure 12** permet d'explorer l'ensemble des données et de visualiser de manière interactive l'interdépendance entre les données via différents types de visualisations (tableaux, diagrammes, corrélations, histogrammes…).

Le deuxième onglet (**figure 13**) permet d'accéder au modèle de ML entraîné sur l'ensemble des données. L'utilisateur peut ainsi prédire de nouvelles données en ajustant les données d'entrée ou encore obtenir des indications sur la manière dont le modèle produit un résultat. Cela permet de comprendre comment chaque paramètre influe sur le résultat et, éventuellement, de mieux appréhender le procédé. La partie de gauche permet d'expliquer les caractéristiques du modèle dans sa globalité avec l'importance de chaque paramètre. La partie de droite permet de prédire de nouvelles sorties à partir d'un jeu de données d'entrée spécifié par l'utilisateur ainsi que l'explication de la prédiction à partir de ce jeu de données.

Conclusion & Perspectives

Les modèles de ML et l'explicabilité associée peuvent être d'une grande aide à la maîtrise des procédés industriels. L'accès aux chemins de décision des algorithmes, en expliquant quelles caractéristiques ont influencé une décision particulière permettant de faciliter l'intégration de ces méthodes dans un environnement industriel.

En effet, dans des métiers où l'expertise et le savoir-faire sont indispensables au développement et à la maîtrise des procédés, l'intégration de modèles de type « boîte noire » peut être reçue avec méfiance. Afin de surmonter cela, le développement de modèles de ML experts spécifiques à un secteur, développés en étant capable d'expliquer la manière dont sont exploitées les données, en accord avec l'expertise humaine et l'homme du métier, ouvre des perspectives prometteuses. Une telle approche favorise une intégration efficace, avec confiance et de manière raisonnée, de l'« intelligence augmentée », alignée avec les compétences et besoins réels de l'industrie. Contrairement aux grands modèles de langage généralistes en développement, l'usage de l'IA dans un environnement industriel, qui plus est des procédés spéciaux, se doit d'être spécifique et ciblé. Il est essentiel d'intégrer la connaissance propre à chaque secteur, enrichie non seulement par les données, mais également l'expertise humaine associée. ■

Références

[1] L.Julia ; *L'intelligence artificielle n'existe pas* (préf. Jean-Louis Gassée), Paris, Éditions First, 2019, 200 p. (ISBN 978-2-412-04340-0).

[2] Y. Le Cun ; *L'invité de 7h50*, France Inter, 12 avril 2023, interview radio.