

Module 7 – Conception d'un POC Synthèse et points clés

- 01** L'objectif principal est de détailler les étapes nécessaires **pour parvenir à un POC (proof of concept) en IA, c'est-à-dire concevoir un algorithme, en identifiant un cas d'usage spécifique et en sélectionnant des variables prédictives pertinentes.**
- 02** Les données brutes nécessitent un **prétraitement (pre-processing) avant d'être utilisées par un algorithme.** Cela comprend la gestion des données manquantes, où les valeurs manquantes sont souvent remplacées par la moyenne des valeurs de la variable.
- 03** Les variables catégorielles ou qualitatives comme la profession ou le pays sont transformées en variables quantitatives pour être comprises par un algorithme. Chaque modalité de la variable catégorielle est considérée comme une variable distincte.
- 04** Pour résoudre les problèmes de différence d'échelle entre les variables, **une mise à l'échelle est effectuée pour ramener toutes les variables dans la même plage de valeurs.**
- 05** Pour concevoir et tester l'algorithme, la base de données est divisée en deux échantillons : **un échantillon d'entraînement pour construire l'algorithme et un échantillon de test pour évaluer ses performances.** La validation croisée est une méthode plus sophistiquée impliquant plusieurs fichiers de test.
- 06** Différents algorithmes sont testés, commençant souvent par la régression logistique pour les données structurées. Si les performances ne sont pas satisfaisantes, d'autres algorithmes plus complexes tels que les SVM (Systèmes à Vaste Marge), les réseaux de neurones, les forêts aléatoires ou le Boosting Gradient sont envisagés.
- 07** Pour évaluer les performances d'un algorithme de classification (les plus fréquents), la matrice de confusion est utilisée. Elle permet de distinguer les vrais positifs, vrais négatifs, faux positifs et faux négatifs.
- 08** Un faux positif est le cas où l'algorithme a prédit quelque chose qui n'arrive pas, comme par exemple qu'un client va acheter un produit alors que dans la réalité non. Un faux négatif, c'est le contraire. C'est le cas où l'algorithme a prédit que rien n'arriverait alors que si.
- 09** L'erreur la plus grave est généralement celle des faux négatifs, car elle peut entraîner des conséquences coûteuses. Les indicateurs de performance, tels que l'accuracy et le taux d'erreur, sont utilisés pour évaluer l'algorithme.
- 10** Une fois l'algorithme construit, testé et évalué, il peut être mis en production pour une utilisation à grande échelle au sein de l'entreprise.