

Évaluez le niveau de maturité de votre entreprise en matière de valorisation des données grâce à notre

# **CARTE DE PERFORMANCE DES DONNÉES**

**Cadre de maturité pour la valorisation  
des données et l'intelligence artificielle**

Un guide en 7 phases pour bâtir  
une pratique de science des  
données durable et performante

# PHASE 1



Avant de vous lancer dans le monde de la **Data Science et du Machine Learning**, il est essentiel d'évaluer les actifs existants, de comprendre les faiblesses de votre organisation, d'identifier les sources potentielles de valeur à travers les projets, etc. **Évaluer votre niveau de maturité** en matière de valorisation des données au sein de votre entreprise vous aidera à savoir où concentrer vos efforts, que ce soit avant ou pendant vos **investissements** en Data Science et Machine Learning.

Cet outil **interactif** vous attribuera un score global basé sur vos réponses, accompagné de commentaires sur les points à **améliorer** et les prochaines étapes à envisager. Gardez à l'esprit qu'il s'agit d'un outil qualitatif destiné à orienter la prise de décision. Les questions sont réparties en trois catégories : "Données", "Data Science" et "Business & Culture".

Vos actifs de données et la manière dont ils sont structurés reflètent votre **niveau de préparation** à créer des solutions complexes en Machine Learning reposant sur ces données. Vos projets ou initiatives actuels en Data Science traduisent l'appétence de votre organisation pour ce type de **solutions**. Enfin, votre culture et l'aspect business indiqueront le niveau de soutien dont vous bénéficierez au fil de votre parcours, ainsi que la maturité organisationnelle de vos équipes face à la Data Science et au Machine Learning.



## Pourquoi ce cadre?

L'intelligence artificielle et la science des données ont le pouvoir de **transformer profondément** les organisations, à condition que leur mise en œuvre soit **progressive**, structurée et alignée sur les priorités d'affaires. Ce cadre méthodologique a été conçu pour accompagner les entreprises tout au long de leur parcours, de **l'évaluation de la maturité data** jusqu'au déploiement de projets IA en production.

Le cadre se décline en **7 phases complémentaires**, chacune adressant un jalon clé de votre transformation:

- **Évaluer** la maturité de vos données à l'aide de la Data Performance Score Card
- **Identifier** des cas d'usage IA à fort potentiel grâce au cadre T.L.D (Temps – Logique – Données)
- **Valider** la faisabilité technique des projets IA : disponibilité des données, existence de cas comparables, contraintes de déploiement
- **Prioriser** les projets IA selon leur valeur ajoutée et leur faisabilité (matrice de décision)
- **Définir** un périmètre clair pour chaque projet (Scope of Work)
- **Constituer** une équipe IA adaptée aux besoins de chaque projet (Data Scientists, MLOps, ingénieurs de données, etc.)
- **Renforcer** les chances de succès par une bonne gouvernance, des objectifs clairs, une culture data forte et un environnement MLOps robuste



Chaque phase est accompagnée d'outils concrets, de **canevas décisionnels**, de profils à mobiliser et de bonnes pratiques organisationnelles. Ce framework est le fruit **d'années de projets** menés dans divers secteurs (énergie, télécoms, services publics, finance), et s'appuie sur une approche à la fois **technique, stratégique et humaine**.

Il s'adresse autant aux organisations en phase d'exploration qu'à celles prêtes à industrialiser leurs démarches en IA. Utilisé de manière progressive, il vous aidera à structurer une pratique de science des données alignée, **durable et mesurable**, tout en favorisant l'engagement des parties prenantes internes.

# QUESTIONS SUR LES DONNÉES

- 1 Disposez-vous d'une traçabilité des données pour identifier vos sources de données et comprendre les flux de données ?

*\*\*La traçabilité des données (ou data lineage) est le processus qui consiste à comprendre, enregistrer et visualiser les données à mesure qu'elles circulent depuis les sources jusqu'à leur utilisation finale.*

- 2 Disposez-vous d'un outil adapté pour le nettoyage et la transformation des données (outil ETL)?

- 3 Avez-vous une pratique d'Intelligence d'affaires avec des outils de reporting et des ressources dédiées?

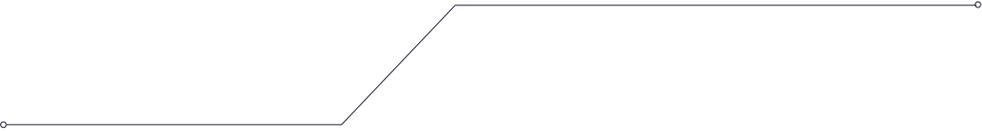
- 4 Disposez-vous d'un entrepôt de données (data warehouse)?

- 5 Si vous avez une pratique d'Intelligence d'Affaires, à quel point la jugez-vous réussie sur une échelle de 0 à 10?

0 signifiant pas du tout réussie, et 10 une pratique extrêmement réussie.

- 6 Si vous avez une pratique d'Intelligence Artificielle, depuis combien de temps est-elle en place?



- 
- 7 Avez-vous un programme de gouvernance des données en place qui définit qui peut agir sur les données?

## QUESTIONS SUR LA DATA SCIENCE

- 8 Avez-vous réalisé des preuves de concept en IA ?
- 9 Avez-vous acheté une solution alimentée par de l'IA (y compris des outils en libre-service)?
- 10 Combien de projets en IA avez-vous identifiés dans votre organisation?
- 11 Avez-vous un analyste de données dans l'entreprise qui réalise vos analyses de données?

## QUESTIONS SUR LES AFFAIRES ET LA CULTURE

- 12 Qualifieriez-vous votre entreprise de 'data-driven' (axée sur les données)?

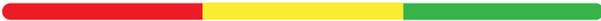


- 13 La direction générale soutient-elle les futurs investissements en Intelligence Artificielle ?
- 14 Bénéficiez-vous du soutien de l'équipe IT dans votre processus?
- 15 Y a-t-il des départements ou unités internes qui ont un intérêt pour les solutions en Intelligence Artificielle?





# Matrice des résultats



Inférieur 15

15-25

26+

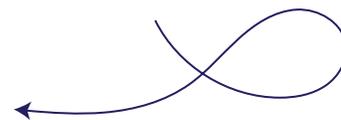


Si score  
inférieur à  
15

Devenir une organisation axée sur les données est un processus difficile et mettre en place une culture où la valeur est créée à travers les données est un long chemin. Votre niveau de maturité reflète la nécessité d'investir dans le développement d'une pratique plus forte en intelligence d'affaires avec les outils appropriés. Obtenir le soutien de la direction générale vous aidera également à gagner en traction au sein de votre organisation. Notre modèle a évalué votre organisation comme étant environ à 25% de maturité en matière de valorisation des données. Ne baissez pas les bras! Votre parcours de données en est encore à ses débuts, et vous avez de la place pour progresser!

Voici quelques actions concrètes que vous pouvez entreprendre pour commencer à améliorer votre niveau de maturité:

- Commencez par optimiser votre pratique de la Business Intelligence, ou créez-en une si ce n'est pas encore le cas.
- Assurez-vous que les besoins de vos utilisateurs soient satisfaits de manière appropriée.
- Investissez dans des outils de nettoyage des données, de traçabilité des données (data lineage) et dans un entrepôt de données, si ce n'est pas déjà le cas.
- Commencez à réfléchir à la gouvernance des données au sein de votre organisation afin de mieux suivre les données et sécuriser leur accès.
- Identifiez des projets potentiels basés sur les données à travers votre organisation.
- Rapprochez-vous des lignes de métier (LoBs) et essayez d'identifier et de positionner des projets basés sur les données.
- Créez des ateliers internes, des webinaires ou des sessions de démonstration au sein de votre organisation. Cela soutiendra la transition vers une organisation axée sur les données et encouragera les collaborateurs à réfléchir à des façons d'utiliser les données.
- Assistez à des conférences et à des webinaires pour rencontrer d'autres leaders comme vous et apprendre de leurs succès et de leurs échecs.





# Matrice des résultats



Inférieur 15

15-25

26+

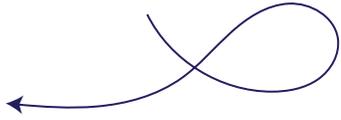


Si score entre  
15 - 25

Votre organisation semble disposer de bons atouts en matière de valorisation des données. Nous pouvons certainement améliorer ce processus en investissant dans des outils appropriés et en créant des bonnes pratiques. Vous avez probablement une direction générale supportive et une bonne écoute de la part de différents départements et lignes de produits. Votre parcours en Data Science et Machine Learning peut commencer sur de bonnes bases si nous optimisons certains de vos actifs existants. Notre modèle a évalué votre organisation comme étant à environ 50% de maturité en matière de valorisation des données. Continuez ainsi! Votre parcours en Data Science et Machine Learning pourrait commencer dès aujourd'hui, mais optimisons vos actifs existants pour vous assurer de partir sur des bases solides!

Voici quelques actions concrètes que vous pouvez entreprendre pour commencer à améliorer votre niveau de maturité:

- Investissez plus sérieusement dans la gouvernance des données pour vous aider à prendre le contrôle sur le flux de vos données.
- Passez votre pratique de la Business Intelligence d'une approche 'demande et exécution' à une approche plus proactive, où vous anticipez les besoins de vos utilisateurs et créez des travaux reproductibles (tableaux de bord, demandes de données et rapports).
- Investissez davantage dans une équipe de data engineering qui nettoiera vos données et vous aidera à disposer de données plus fiables et propres, tant pour la Business Intelligence que pour la Data Science.
- Créez des ateliers internes, des webinaires ou des sessions de démonstration au sein de votre organisation. Cela soutiendra la transition vers une organisation pleinement axée sur les données et encouragera les collaborateurs à réfléchir à des façons d'utiliser les données.
- Si ce n'est pas déjà le cas, commencez à identifier des projets en Data Science en utilisant notre cadre (2—Identifiez les projets à forte valeur ajoutée pour votre organisation).



COMMENCEZ VOTRE PARCOURS

# Matrice des résultats



Si Score au-delà de 26

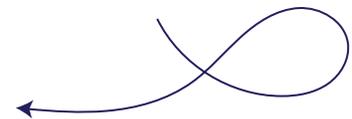
Votre organisation affiche une grande maturité en matière de valorisation des données. Il semble que vous soyez en contrôle de la situation. Vous bénéficiez d'un bon soutien interne et votre équipe BI fait un excellent travail. Les chances de succès en Data Science et Machine Learning sont élevées.

Notre modèle a évalué votre organisation à au moins 75% de maturité en matière de valorisation des données. Vous êtes définitivement prêt pour un grand saut dans la Data Science et le Machine Learning!

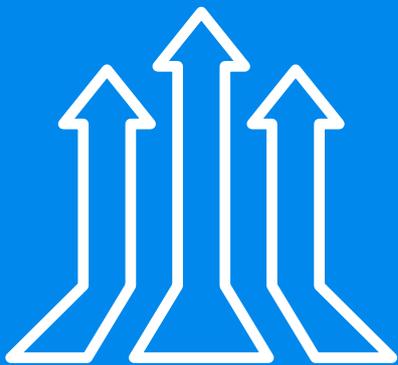
Voici quelques actions concrètes que vous pouvez entreprendre pour consolider vos forces existantes:

- Optimisez votre gouvernance des données pour vous aider à prendre le contrôle sur le flux de vos données.
- Créez des ateliers internes, des webinaires ou des sessions de démonstration au sein de votre organisation. Cela soutiendra la transition vers une organisation entièrement axée sur les données et incitera les collaborateurs à réfléchir à des moyens d'utiliser les données.
- Si ce n'est pas déjà le cas, commencez à identifier des projets en Data Science en utilisant notre cadre (2—Identifiez les projets à forte valeur ajoutée pour votre organisation).

COMMENCEZ VOTRE PARCOURS



# PHASE 2



## **Identifier des projets à forte valeur ajoutée pour votre organisation**

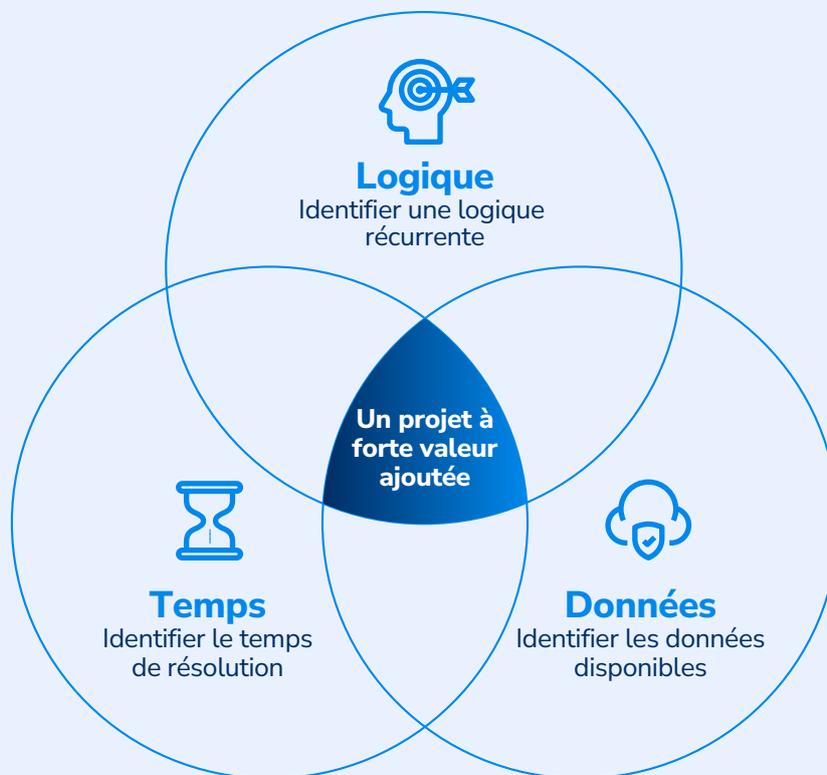
Dans la phase 1, nous avons évalué le niveau de maturité pour la valorisation des données. Maintenant, nous avons suffisamment d'informations pour améliorer notre fondation et la gestion des données afin de nous préparer aux projets de Data Science et de Machine Learning.

Détecter des projets à forte valeur ajoutée au sein d'une organisation peut sembler facile. En fonction du type de données disponibles, vous pourriez repérer des projets ici et là. Mais comment savoir si un projet aura une forte valeur pour l'organisation ?

Notre cadre pour détecter les projets à forte valeur vous aidera à sélectionner les projets, ce qui facilitera la construction d'une feuille de route pour votre pratique de Data Science. Ce cadre repose sur trois dimensions qui permettent de filtrer les projets.

## Notre cadre T.L.D. :

Le cadre T.L.D. (Temps, Logique et Données) se concentre sur le processus métier sous-jacent d'un projet spécifique afin de déterminer si celui-ci est adapté au Machine Learning. De plus, il permet d'évaluer si un système de Machine Learning contribuera à optimiser le processus métier visé par le projet.



## Quel est le temps de résolution du processus métier sous-jacent ?

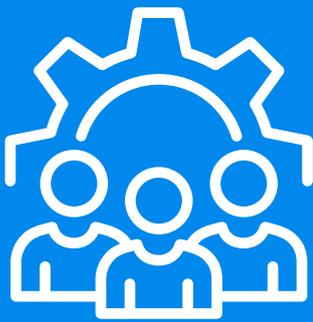
Le temps de résolution définit combien de temps il faut pour accomplir un processus métier ou un sous-processus. Plus ce temps est élevé, plus il est pertinent d'utiliser le Machine Learning pour résoudre le problème lié à ce processus. Par exemple, dans une usine, le processus de lecture d'un code-barres sur une boîte ou un article est réalisé via un système de scan (non alimenté par du Machine Learning). Ce processus est rapide car le temps de lecture du code-barres est très court. Cela signifie que le temps de résolution de ce processus est très faible, et donc que l'usage du Machine Learning y apporte peu de valeur en tant que projet.



## **Y a-t-il une récurrence de la logique dans la manière de résoudre le processus métier sous-jacent?**

La récurrence de la logique est un élément essentiel du cadre T.L.D. Elle représente à quel point la résolution des problèmes dans un processus métier donné est prévisible. Elle reflète l'existence de schémas et de logiques que l'on peut retrouver lors de la résolution d'un problème. Par exemple, lorsque vous essayez de déterminer si un candidat dans votre organisation deviendra un bon employé, vous suivez une certaine logique et identifiez des schémas à partir de vos expériences précédentes avec d'autres candidats pour prédire son potentiel. Lorsqu'il existe des schémas riches dans la manière de résoudre un processus métier, cela signifie que le projet est particulièrement adapté au Machine Learning.

La récurrence de la logique est présente dans presque tous les processus, mais certains en présentent des schémas plus riches et plus clairs que d'autres.



## **Des données sont-elles générées pour refléter le processus métier ?**

La dernière dimension est celle des données. Les données n'ont pas nécessairement besoin d'être déjà numériques ou structurées. Elles peuvent être non structurées (texte, image, etc.). Par exemple, la création d'un assistant virtuel nécessiterait des données historiques de conversations, qui sont sous forme de texte brut.

Une grande quantité de données n'est pas toujours nécessaire : cela dépend de la complexité du problème que l'on cherche à résoudre. Certains problèmes exigent un grand volume de données, d'autres non.

Ce qui est essentiel, c'est d'avoir commencé à collecter des données. Cela augmentera considérablement la valeur d'un projet en Machine Learning.

# EXEMPLES

## PRÉDICTION DU MARCHÉ BOURSIER

Chemin	Score	Commentaires
Temps de résolution	Élevé 	Lorsqu'un humain essaie de prédire le marché boursier, il effectue des recherches, consulte d'autres experts, etc., afin de prendre une décision efficace. Cela peut être un processus long.
Réurrence de la logique	Moyen / Faible 	Parfois, les mouvements du marché boursier peuvent être difficiles à comprendre. Ainsi, prédire l'avenir peut reposer sur une logique obscure (il existe bien une logique, mais elle est difficile à saisir).
Données collectées	Élevé 	De nombreuses sources de données sont disponibles, comme les prix historiques, les actualités, les communications et commentaires des personnes, les états financiers des entreprises, etc.

## ASSISTANT JURIDIQUE EN INTELLIGENCE ARTIFICIELLE

Chemin	Score	Commentaires
Temps de résolution	Élevé 	Lorsqu'un assistant juridique humain aide un avocat à constituer un dossier, ce processus peut prendre du temps, notamment pour effectuer des recherches et lire les informations relatives au cas.
Réurrence de la logique	Élevé 	La logique suivie lors des recherches est récurrente, car l'assistant juridique utilisera son expérience pour savoir ce qu'il/elle doit rechercher.
Données collectées	Élevé 	Le droit est un domaine riche en données provenant de diverses sources : anciens cas, livres de droit, etc.

# EXEMPLES

## PRÉDIRE LE COMPORTEMENT D'UN CLIENT

Chemin	Score	Commentaires
Temps de résolution	Élevé 	Pour qu'un humain analyse et soit capable de prédire le comportement d'un client, cela nécessite une compréhension approfondie du client et du produit, ainsi qu'une expérience préalable en marketing client. Cela rend le temps de résolution très élevé.
Réurrence de la logique	Élevé 	La logique qui régit le comportement d'un client peut être capturée et est claire. Un client a tendance à suivre des schémas spécifiques.
Données collectées	Moyen / Faible 	Prédire le comportement d'un client nécessite des données sur le comportement passé du client, mais aussi un profil complet du client. Les profils des clients (y compris les données démographiques) sont difficiles à collecter pour de nombreuses entreprises.

**L'étape suivante consiste à télécharger le cadre: Identifier les projets de Machine Learning à forte valeur ajoutée pour votre organisation.**

Vous cherchez du soutien dans votre parcours?

**CONTACTEZ-NOUS DÈS MAINTENANT**

# PHASE 3



## La faisabilité d'un projet de Machine Learning

Maintenant que nous savons comment détecter des projets à fort potentiel, nous pouvons passer à l'étape suivante et apprendre à évaluer la faisabilité d'un projet.

Il est important de noter qu'un projet à fort potentiel ne signifie pas nécessairement qu'il est faisable. La faisabilité signifie qu'un projet est techniquement réalisable. En général, cela nécessite des connaissances techniques approfondies ainsi qu'une bonne expérience de l'industrie pour savoir si un projet est vraiment faisable. Et en raison de l'incertitude inhérente au Machine Learning, un projet faisable sur le papier peut devenir irréalisable lors du développement.

Ce cadre a pour objectif de simplifier votre décision concernant la faisabilité d'un projet. Nous recommandons de valider la faisabilité auprès d'un spécialiste en Machine Learning qui vous guidera dans l'évaluation de la faisabilité de vos projets.



## Question à se poser lors de l'analyse de faisabilité

Les données sont-elles disponibles en interne ou doivent-elles être acquises en externe ?

Très souvent, ce qui limite un projet, c'est la disponibilité des données. Les entreprises et les dirigeants découvrent que les données internes ne suffisent pas à mener un projet spécifique, qu'aucun tiers ne peut fournir ces données, ou encore que l'achat de données auprès de prestataires externes est trop coûteux.

Le cas d'usage le plus courant se trouve dans le marketing, où les entreprises souhaitent lancer des projets de Machine Learning sur leurs données clients (comme la prédiction du comportement client), et réalisent qu'il leur manque des données essentielles sur leurs clients (âge, sexe, données socio-démographiques, etc.).

Lorsqu'elles se tournent vers des fournisseurs tiers, cela peut devenir très coûteux, en particulier lorsqu'il s'agit de millions de clients avec de nombreuses caractéristiques souhaitées.

Cela reflète la capacité de l'entreprise à collecter les données nécessaires à un projet spécifique. Parfois, la collecte des données n'est pas possible pour des raisons marketing ou éthiques, et parfois cela peut être long, car il faut du temps pour accumuler une quantité suffisante de données.

**Ainsi, la disponibilité des données peut jouer un rôle crucial dans l'analyse de faisabilité.**

## **Le projet dispose-t-il de cas d'usage existants sur le web ou dans d'autres entreprises?**

Si un projet est déjà documenté sur le web comme cas d'usage, cela signifie probablement qu'il est faisable. Par exemple, si vous recherchez "prédiction de l'attrition client" sur Google, vous trouverez de nombreux cas d'usage ainsi que des entreprises expliquant comment elles ont mis en œuvre ce projet de Machine Learning dans leur organisation.

En assistant à des conférences dans votre secteur ou sur l'IA, vous pourriez également entendre parler de cas d'usage similaires à vos projets potentiels, expliquant comment ceux-ci ont été implémentés.

Tous ces éléments sont des indicateurs positifs de faisabilité. Toutefois, cela ne garantit pas que votre projet sera techniquement faisable en pratique, car chaque entreprise et chaque projet sont uniques. Cela reste cependant un bon indicateur.

## **À quel point serait-il difficile de déployer et de maintenir le modèle en production?**

Un autre facteur à considérer dans l'analyse de faisabilité est la manière dont le modèle sera déployé, et le coût de sa maintenance en production.

Par exemple, une entreprise de télécommunications souhaite créer un modèle de Machine Learning capable de détecter toute anomalie dans ses données réseau. Ce projet nécessite la création de pipelines d'ingestion en production. Ces pipelines sont ensuite connectés au modèle de Machine Learning déployé.

Créer de tels flux de données pour l'ensemble du réseau est très coûteux en infrastructure matérielle, et les flux nécessitent aussi des efforts de maintenance et de mise à jour.

Déterminer la complexité du déploiement et de la maintenance peut devenir très technique, et nécessiter l'avis d'un spécialiste (ingénieur en Machine Learning ou expert MLOps).



## Scénarios

Une entreprise de vente au détail souhaite analyser si un projet de prédiction de l'attrition client est faisable. L'entreprise a évalué que ce projet de prédiction de l'attrition client présente une grande valeur.

Tout d'abord, l'entreprise a examiné les données collectées et disponibles. Elle a constaté qu'elle dispose de données transactionnelles sur les clients, ainsi que d'informations de base sur leurs clients, comme l'adresse. D'autres données pourraient être disponibles, notamment les données sur les appels au service client, y compris la durée des appels, le nombre d'appels par mois, etc. Avoir accès à des données socio-démographiques supplémentaires aurait été un atout pour soutenir le projet, mais l'entreprise ne souhaite pas demander ces informations à ses clients et ne veut pas les acheter auprès de fournisseurs tiers. Cela rend donc le projet techniquement faisable du point de vue des données.

L'un des responsables a décidé de faire une recherche sur Google en tapant "prédiction de l'attrition client" et a constaté que la prédiction de l'attrition est un projet bien documenté, avec de nombreux cas d'usage disponibles en ligne. Cela rend le projet faisable d'un point de vue historique des cas d'usage.

Du point de vue du déploiement et de la production, les modèles pourraient être déployés sous forme de script simple contre les bases de données, et cela fonctionnerait assez bien. D'autres déploiements plus sophistiqués sont possibles, comme le déploiement sous forme de microservice, ce qui permettrait l'intégration avec d'autres outils et logiciels. Cela rend le projet faisable d'un point de vue du déploiement avec une complexité moyenne à faible.

## Performance globale

Données	Assez pour commencer le premier POC/Prototype
Suivi des cas d'usage	Haute disponibilité
Déploiement et maintenance en production	Complexité moyenne à faible

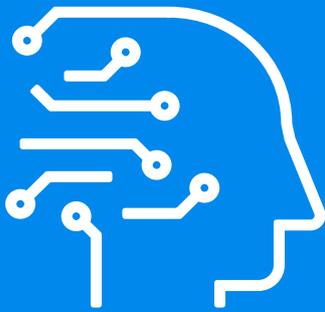
## Que faire si le projet n'est pas vraiment faisable ?

Un projet qualifié de non faisable peut tout de même être mené en fonction de l'aspect qui pose problème. Si le projet est non faisable en raison d'un manque de données, le projet peut être retardé en attendant de collecter suffisamment de données. Si le projet est non faisable en raison d'un manque de cas d'usage (aucun cas existant trouvé), le projet peut être mené en mode R&D. Si un projet n'est pas faisable du point de vue de la production, l'entreprise peut commencer à le mettre en œuvre à une échelle réduite ou dans un environnement de simulation. Travailler avec des Architectes en Machine Learning peut aider à concevoir un environnement optimisé et conçu pour la production, ce qui réduira les coûts et les efforts de production.

Vous cherchez du soutien dans votre parcours?

**CONTACTEZ-NOUS DÈS MAINTENANT**

# PHASE 4



## **Cadre de priorisation des projets d'apprentissage automatique**

Maintenant que nous avons appris comment évaluer la faisabilité d'un projet, vous avez probablement évalué de nombreux projets. Comment pouvez-vous prioriser tous ces projets et décider lesquels devraient être vos priorités à court terme dans votre pratique de la Science des Données?

Dans les phases précédentes, nous avons évalué la faisabilité et la valeur d'un projet. Ces deux facteurs deviendront nos vecteurs de décision pour la priorisation des projets. Nous utiliserons des quadrants qui croisent la faisabilité avec la valeur pour comprendre quels projets devraient être une priorité.

## Quadrants Valeur/Faisabilité



Chaque projet doit être ajouté au quadrant en fonction de sa valeur et de sa faisabilité. Le carré en vert devrait regrouper tous les projets ayant la faisabilité et la valeur les plus élevées.

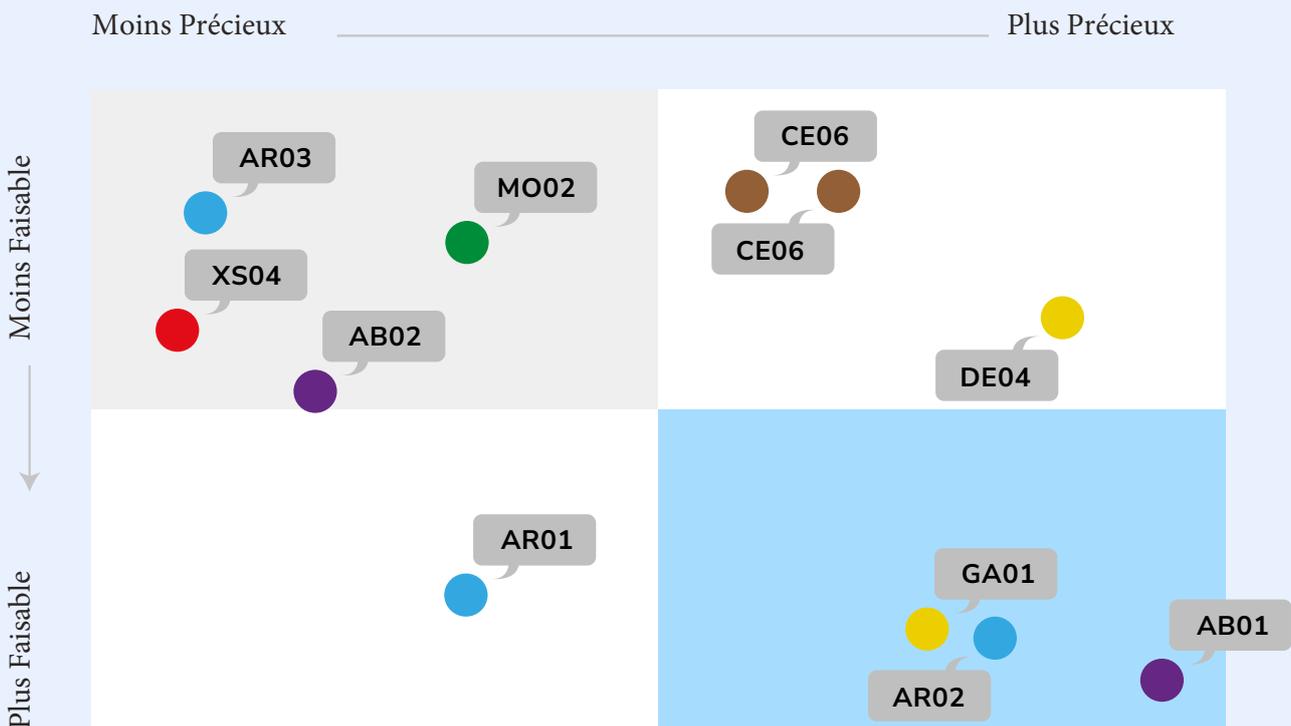
Notez qu'il est important de comparer les projets en termes de faisabilité et de valeur. Un projet est toujours précieux et réalisable par rapport à d'autres projets.

Il s'agit d'une approche qualitative, nous vous conseillons donc de mener cette analyse avec les membres de votre équipe, car de multiples opinions augmenteront vos chances d'être précis.

Un petit conseil est de regrouper les projets en catégories (comme l'augmentation des ventes, le contenu marketing, l'expérience client, etc.). Si vous avez un grand nombre de projets détectés, vous pouvez les diviser en quadrants plus petits par catégorie avant d'exécuter un quadrant final.

## Exemples

Après avoir évalué la valeur (cadre 2) et la faisabilité (cadre 3), vous pouvez créer votre quadrant en comparant la faisabilité et la valeur entre les projets. Dans l'exemple suivant, 3 projets se trouvent dans la zone idéale et deviendront donc la priorité de votre feuille de route à court terme. Gardez à l'esprit que les autres projets ne sont pas disqualifiés ni laissés de côté. Ils peuvent être ajoutés à la feuille de route à moyen/long terme en fonction de leur ordre de priorité.



Vous cherchez du soutien dans votre parcours?

**CONTACTEZ-NOUS DÈS MAINTENANT**

# PHASE 5



## Étendue des Travaux (EDT)

Après avoir priorisé les projets, nous connaissons les projets avec lesquels nous devons commencer. Pour lancer notre première initiative, nous devons définir l'étendue des travaux pour chaque projet.

# Étendue des Travaux (EDT)

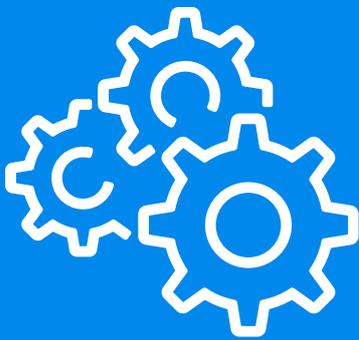
Le document suivant est un modèle pour rédiger l'étendue des travaux.

Chemin	Détails	Commentaires
Contexte commercial	Fournissez une introduction sur le processus métier sous-jacent au projet et sur la manière dont il aidera l'entreprise.	Le contexte commercial aidera l'équipe à garder à l'esprit les objectifs du projet.
Problématique	Quel problème essayons-nous de résoudre?	Le problème doit être bien défini pour éviter de mener à des résultats erronés.
Aperçu technique	Explication de ce qui est techniquement requis et des types de techniques nécessaires (apprentissage supervisé? Classification? Vision par ordinateur?)	Doit être réalisé avec l'équipe technique (un data scientist ou un ingénieur en apprentissage automatique).
Données requises	Définir quelles sources de données sont nécessaires pour réaliser le projet. Détailler également les variables disponibles avec leur type de données.	Doit être détaillé avec la source de données.
Hypothèses	Toute hypothèse qui peut impacter la livraison.	Le type d'hypothèses peut être la collaboration d'une ressource spécifique extérieure à votre équipe. Ou que le projet sera mené en mode Recherche, etc.
Livrables	Qu'est-ce qui est attendu comme résultat ?	Devons-nous construire un logiciel? Ou devons-nous livrer un rapport ? Présenter nos résultats?
Effort estimé	Combien de temps est nécessaire?	Peut être en heures ou en jours.
Expertise requise	Quels types de compétences sont nécessaires ? Et quels types de ressources ?	Avons-nous besoin de compétences en vision par ordinateur ? Ou d'une bonne connaissance de SQL ? Avons-nous besoin de développeurs de logiciels pour intégrer le modèle dans un logiciel ?
Exclusions	Qu'est-ce qui est exclu du projet?	Nous pouvons exclure le déploiement s'il ne fait pas partie du périmètre et qu'il sera réalisé ultérieurement. Ou exclure une phase d'analyse de données spécifique, etc

Vous cherchez du soutien dans votre parcours?

**CONTACTEZ-NOUS DÈS MAINTENANT**

# PHASE 6



## **Profils nécessaires pour construire votre équipe de Science des Données**

Maintenant que nous savons quels sont les projets prioritaires, nous avons défini le périmètre de nos projets et nous pouvons passer à la constitution de l'équipe et au recrutement de talents. Le monde de la science des données est complexe et il existe sur le marché bien plus que de simples data scientists. De nombreux profils différents sont disponibles, et une équipe appropriée est composée de profils complémentaires issus de différents horizons.

Votre capacité à identifier les profils nécessaires et à comprendre les compétences requises pour mener à bien les projets que vous avez est très importante. Un projet de R&D ne nécessite pas les mêmes profils que des projets axés sur le business. De plus, le déploiement de modèles et la gestion du cycle de vie de l'apprentissage automatique requièrent des compétences spécifiques. Nous allons commencer par identifier les différents rôles disponibles sur le marché.

# Les différents rôles possibles en science des données.

Titre	Rôle	Compétences	Outils
Analyste de données	Analyser les données et traiter les demandes de données des clients internes.	Analyse commerciale / Ingénierie des données (SQL principalement) / Reporting / Statistiques	SQL / Tableau / Excel / PowerBI / R / SPSS
Analyste métier	Comprendre les besoins métiers pour chaque projet. Traiter les demandes des clients internes.	Analyse commerciale / Reporting / Communication	SQL / Powerpoint / Excel / PowerBI / Tableau
Data Scientist	Développer des modèles d'apprentissage automatique. Effectuer des analyses de données si nécessaire.	Science des données / Statistiques / Apprentissage automatique / Mathématiques / Reporting	Python / R / SQL / Jupyter Notebooks / Tensorflow / Pytorch / Git
Chercheur Scientifique	Rechercher une approche innovante pour résoudre les problèmes d'apprentissage automatique.	Science des données / Apprentissage automatique / Mathématiques / Statistiques	Python / C++ / C / R / SQL / Notebooks / Tensorflow / Pytorch
Ingénieur en Apprentissage Automatique	Créer des solutions d'apprentissage automatique, du développement des modèles au déploiement.	Science des données / Apprentissage automatique / Statistiques / Ingénierie des données / Déploiement de modèles / roservices / Conteneurs	Python / R / SQL / Tensorflow / Pytorch / Git / Cloud providers / Docker / Kubernetes / Go / C# / C++
Ingénieur MLOps / Ingénieur Logiciel ML	Science des données / Apprentissage automatique / Statistiques / Ingénierie des données / Déploiement de modèles / Microservices / Conteneurs	Ingénierie des données / Déploiement de modèles / Microservices / Conteneurs / Ingénierie logicielle / Infrastructure / Pipelines ML	Python / Git / Cloud providers / Docker / Kubernetes / Go / C# / C++ / CI-CD / Microservices architecture / SQL
Ingénieur de données	Rendre les données disponibles pour les data scientists / ingénieurs en apprentissage automatique.	Préparer les données pour la production. Ingénierie des données / ETL / Lac de données / Big data / Reporting	Python / ETL Tools / Hadoop

# Les tâches du cycle de vie de la science des données et les rôles associés.

Tâche	Quotidien type	Rôle associé
Comprendre les besoins des clients	<ul style="list-style-type: none"> <li>-Rencontrer les clients et discuter de leurs projets</li> <li>-Mener des ateliers et des sessions de conception pour comprendre où se situent les projets de valeur.</li> </ul>	Analyste métier Chef de projet science des données Directeur science des données
Définir le périmètre d'un projet et les exigences	<ul style="list-style-type: none"> <li>-Évaluer la faisabilité du projet</li> <li>-Comprendre les ressources nécessaires</li> <li>-Valider le potentiel du projet</li> <li>-Évaluer l'effort</li> </ul>	Analyste métier Chef de projet science des données Directeur science des données
Collecter les données	<ul style="list-style-type: none"> <li>-Créer des pipelines pour l'ingestion des données</li> <li>-Créer un entrepôt de données</li> <li>-Créer des bases de données (data marts)</li> <li>-Créer un lac de données.</li> </ul>	Ingénieur de données Ingénieur logiciel
Nettoyer les données	<ul style="list-style-type: none"> <li>-Scripts/flux ETL</li> <li>-Scripting</li> <li>-Valider la transformation.</li> </ul>	Ingénieur de données Data scientist Analyste de données
Étiqueter les données	<ul style="list-style-type: none"> <li>-Créer un ensemble de données d'entraînement</li> <li>-Utiliser des statistiques pour estimer les étiquettes</li> <li>-Utiliser des données ouvertes.</li> </ul>	Data scientist Chercheur scientifique
Explorer les données	<ul style="list-style-type: none"> <li>-Visualiser les données</li> <li>-Comprendre la distribution des données</li> <li>-Corrélation</li> <li>-Analyse univariée/multivariée.</li> </ul>	Analyste de données Data scientist Chercheur scientifique
Rapporter sur les données et la performance du modèle	<ul style="list-style-type: none"> <li>-Création de rapports et de tableaux de bord sur la qualité des données</li> <li>-Création de rapports et de tableaux de bord sur les sources de données</li> <li>-Création de rapports et de tableaux de bord sur la performance du modèle.</li> </ul>	Analyste de données Ingénieur MLOps (pour des conseils sur la performance du modèle)
Entraîner les modèles d'apprentissage automatique	<ul style="list-style-type: none"> <li>-Entraîner les modèles</li> <li>-Prétraitement des données</li> <li>-Sélection de modèles</li> <li>-Évaluation de modèles.</li> </ul>	Data scientist Chercheur scientifique Ingénieur en apprentissage automatique

Tâche	Quotidien type	Rôle associé
Optimisation de la phase d'entraînement	<ul style="list-style-type: none"> <li>-Exécution en calcul parallèle</li> <li>-Optimisation du matériel</li> <li>-Exécution sur GPU</li> </ul>	Ingénieur MLOps Ingénieur en apprentissage automatique
Création de nouveaux algorithmes d'apprentissage automatique	<ul style="list-style-type: none"> <li>-Développer un nouvel algorithme d'apprentissage automatique</li> <li>-Recherche de nouvelle approche en ML (Machine Learning)</li> </ul>	Chercheur scientifique
Recherche de nouvelles techniques en apprentissage automatique	<ul style="list-style-type: none"> <li>-Lecture d'articles de recherche</li> <li>-Participation à des conférences</li> </ul>	Chercheur scientifique Data scientist
Refactorisation du code d'apprentissage automatique	<ul style="list-style-type: none"> <li>-Préparer le code d'apprentissage automatique pour la production</li> <li>-Optimiser le code</li> <li>-Réécrire le code</li> </ul>	Ingénieur en apprentissage automatique Ingénieur MLOps
Déploiement de modèles d'apprentissage automatique	<ul style="list-style-type: none"> <li>-Déployer des modèles ML en tant que microservices</li> <li>-Déployer des modèles ML dans un registre de modèles</li> <li>-Créer des images Docker</li> <li>-Déployer des scripts et des applications</li> </ul>	Ingénieur en apprentissage automatique Ingénieur MLOps
Création de pipeline de réentraînement	<ul style="list-style-type: none"> <li>-Créer des pipelines automatisés pour le réentraînement des modèles</li> </ul>	Ingénieur en apprentissage automatique Ingénieur MLOps
Gestion de l'environnement d'apprentissage automatique	<ul style="list-style-type: none"> <li>-Création d'environnement de développement</li> <li>-Gestion de l'environnement de déploiement</li> <li>-Formation et accompagnement de l'équipe sur l'utilisation des outils (outils MLOps)</li> </ul>	Ingénieur MLOps
Définition des exigences de production pour les modèles d'apprentissage automatique	<ul style="list-style-type: none"> <li>-Définition des meilleures pratiques en termes de déploiement</li> <li>-Rencontre avec les data scientists pour comprendre les exigences pour les modèles ML</li> </ul>	Ingénieur MLOps Ingénieur en apprentissage automatique
Configuration des environnements (développement ou production)	<ul style="list-style-type: none"> <li>-Créer un environnement pour le développement et pour la production</li> </ul>	Ingénieur MLOps
Créer un environnement CI/CD Définir des stratégies de réentraînement	<ul style="list-style-type: none"> <li>-Création de pipelines et d'outils d'intégration continue</li> <li>-Création de pipelines et d'outils de livraison continue</li> <li>-Définir les meilleures pratiques pour le réentraînement</li> </ul>	Ingénieur MLOps Ingénieur en apprentissage automatique
Configuration de l'infrastructure ML	<ul style="list-style-type: none"> <li>-Configuration du matériel (ressources) pour les projets ML</li> </ul>	Ingénieur MLOps

Vous cherchez du soutien dans votre parcours?

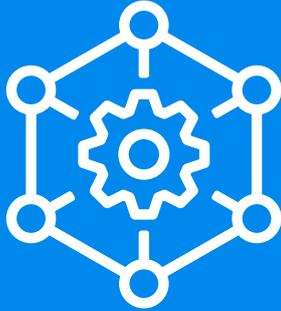
**CONTACTEZ-NOUS DÈS MAINTENANT**

# PHASE 7

## **Clés du succès dans la mise en œuvre d'une pratique de la science des données**

Félicitations ! Vous venez de structurer votre pratique de la Science des Données. Avoir une stratégie basée sur les bons cadres va vous aider à augmenter vos chances de construire une pratique saine. Maintenant, nous allons définir des exigences supplémentaires qui peuvent contribuer à accroître votre succès avec votre pratique.





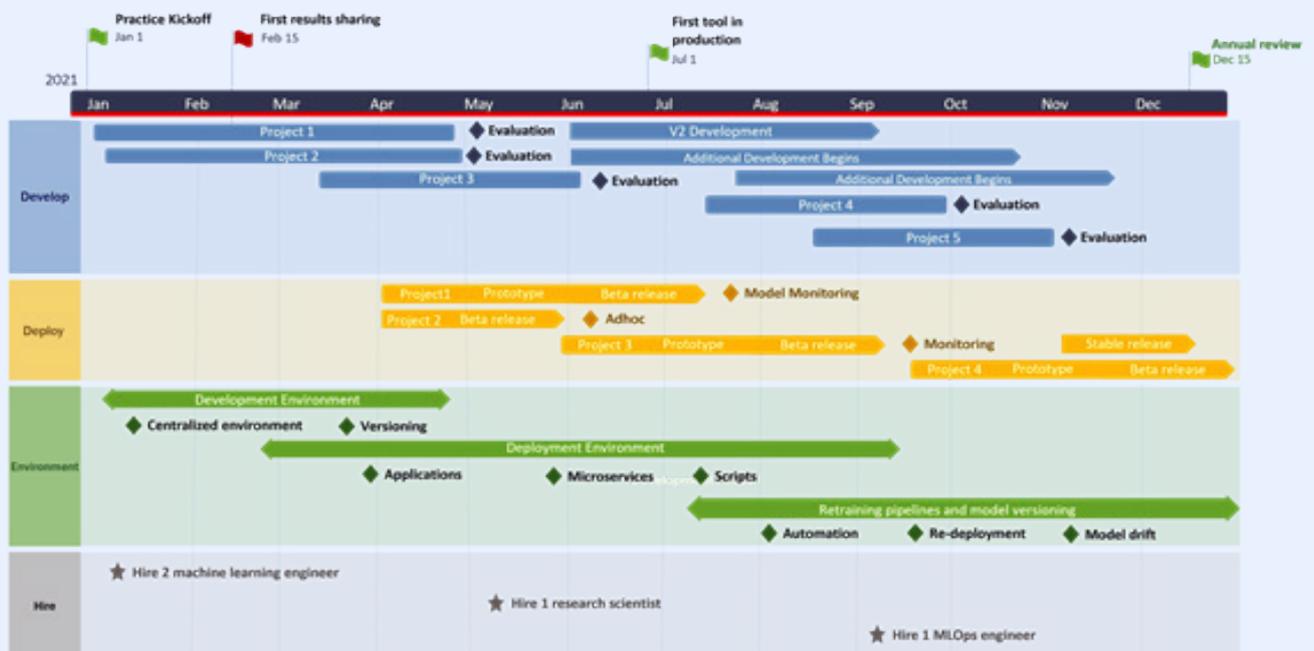
## Définir des objectifs avec une stratégie

Pour valoriser votre pratique, vous devrez définir une vision appropriée qui dictera votre approche et vous aidera à créer des exigences.

La vision peut être définie comme:

- Évoluer vers une organisation axée sur les données
- Monétiser les données/outils créés
- Générer plus de revenus grâce aux outils d'apprentissage automatique
- Réduire les coûts des opérations/du marketing

## Feuille de route





## Environnement MLOps

Il est très important de concevoir un environnement d'apprentissage automatique qui soutiendra notre cycle de développement ML. Cela inclura un environnement de développement durable. Notez que le développement peut être effectué localement sur un ordinateur portable doté de ressources, mais il doit centraliser les métriques et les paramètres de méta et d'expérimentation. De plus, le code doit être validé et centralisé. Idéalement, disposez d'un environnement de développement partagé qui aura un contrôle de version avec un suivi des performances d'entraînement. Ensuite, il est important d'avoir un environnement de déploiement où tous les actifs seront déployés avec une décision concernant la technologie nécessaire au déploiement. L'intégration et la livraison continues (CI/CD) jouent un rôle important dans le passage du développement à la production et doivent être décidées avec soin.

## **Communication**

Il est important de communiquer en interne et en externe autour des solutions créées et des projets innovants. Cela aidera à dynamiser les projets et à créer un écosystème autour des données et de la science des données. Les types de communication peuvent être :

- Participation à des conférences, pour partager des cadres, des succès et des parcours.
- Communiqués de presse autour de nouveaux outils et solutions.
- Communication interne autour des découvertes et des nouveaux projets.
- Devenir un centre de connaissances par le biais de canaux internes et externes.

## **Gestion du changement**

Investir dans la gestion du changement augmentera l'adoption des solutions nouvelles et innovantes créées en interne dans le domaine de la science des données et de l'apprentissage automatique. La gestion du changement contribue à créer une culture axée sur les données au sein de l'entreprise. Cette partie est aussi importante que le développement des solutions.

Pour changer les mentalités et mettre en œuvre des stratégies de gestion du changement, voici quelques conseils:

- Créer des ateliers internes pour partager les connaissances avec les clients internes.
- Organiser des sessions de découverte d'outils avec les clients internes.
- Créer une newsletter interne pour partager les découvertes et les informations.



## Quelques conseils supplémentaires

- Évitez de passer trop de temps dans le développement des projets et développez de manière itérative.
- Assurez-vous de valider les besoins et les hypothèses avec les clients.
- Dans un projet d'apprentissage automatique, assurez-vous d'avoir accès à un expert métier qui aura une compréhension approfondie du secteur ou du domaine où vous appliquez votre apprentissage automatique.
- Adaptez le déploiement aux besoins de votre environnement.
- Surveillez toujours la performance du modèle en production et mettez en place une stratégie de réentraînement.

Vous cherchez du soutien dans votre parcours?

**CONTACTEZ-NOUS  
DÈS MAINTENANT**