

### Qui Suis-Je?

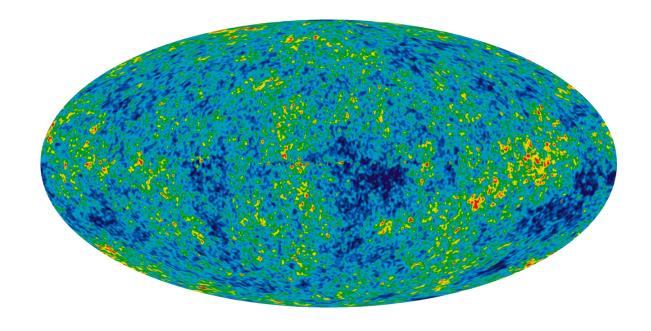


**Ingénieur** 

**PhD Astrophysique** 

**Data Intensive Applications** 

Directeur R&D, associé





#### notre métier



11 personnes | 8 Data Scientists | 2 Phds







angers | paris | la roche sur yon

#### améliorer la performance

Cross Data accompagne les entreprises ambitieuses dans l'intégration de l'Intelligence Artificielle au cœur de leur métier et de leurs process.

A partir de vos données et de nos algorithmes, Cross Data identifie les potentiels de performance cachés dans votre entreprise.

#### concevoir et industrialiser l'IA

Spécialiste en conception et industrialisation de modèles mathématiques, Cross Data développe des solutions IA sur mesure pour ses clients dans l'industrie, l'énergie, les transports, le retail et la santé.

#### associer l'expertise et le pragmatisme

Nous appliquons notre savoir-faire en ML, DL, Vision IA, MLOps, NLP et en mathématiques avec une seule règle : pas de projet sans retour sur investissement.



#### nos modes d'intervention



#### conseil & diagnostic

Accompagnement, Audits Data, Diagnostic IA & Data



#### algorithmes sur mesure intelligence artificielle & maths

Développement spécifique et sur mesure ML, DL, RO



#### formation

Data Science, Python



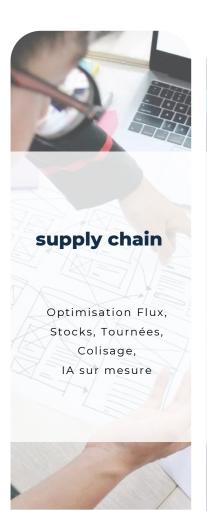




#### nos secteurs de performance













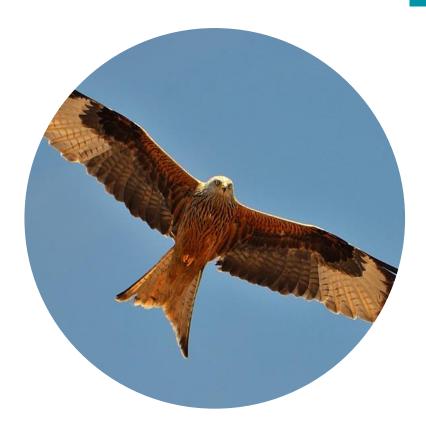


### **Deux parties**

L'IA: panorama en 2023 Un projet d'IA réussi



### IA: définition



Ensemble de théories et de techniques
mises en œuvre en vue de réaliser des
machines capables de simuler l'intelligence
humaine



# IA: définition





### Que sait bien faire une IA?

**Problèmes combinatoires** 

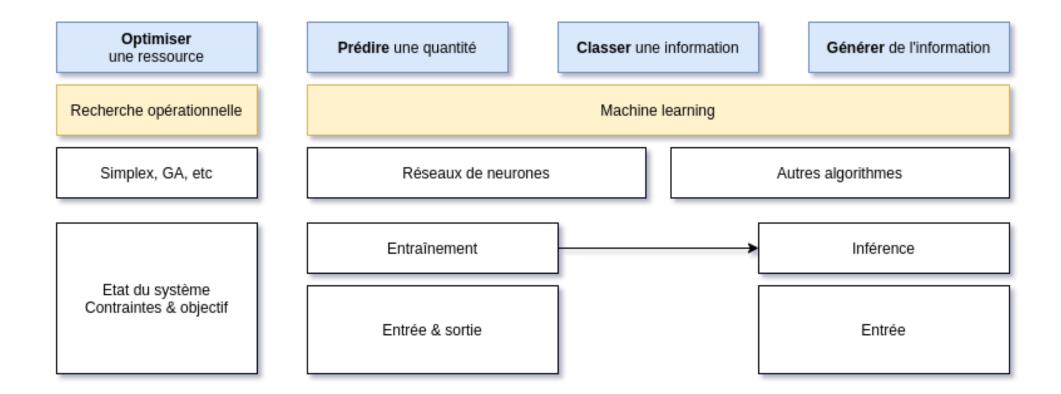
Problèmes avec expertise précise chronophage

Fig. 1.

Les ponts de Kænigsberg en 1759.

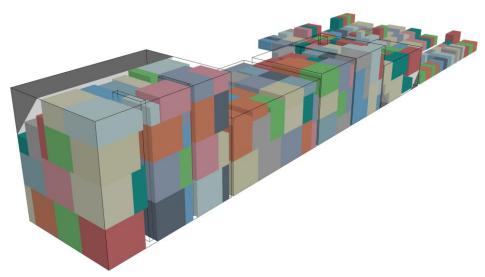


#### **Panorama**





### L'optimisation



1. Maximiser ou minimiser une quantité
2. Avec des contraintes

**Exemples : optimisation de tournées,** colisage, planning etc.



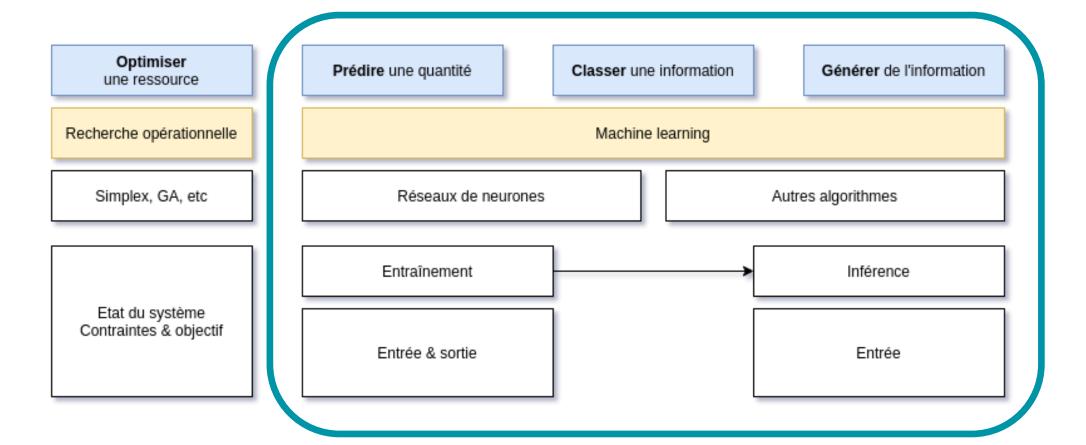
## L'optimisation de planning

Minimiser les délais de production

Solution en moins d'une minute

Cas d'usage

#### **Panorama**



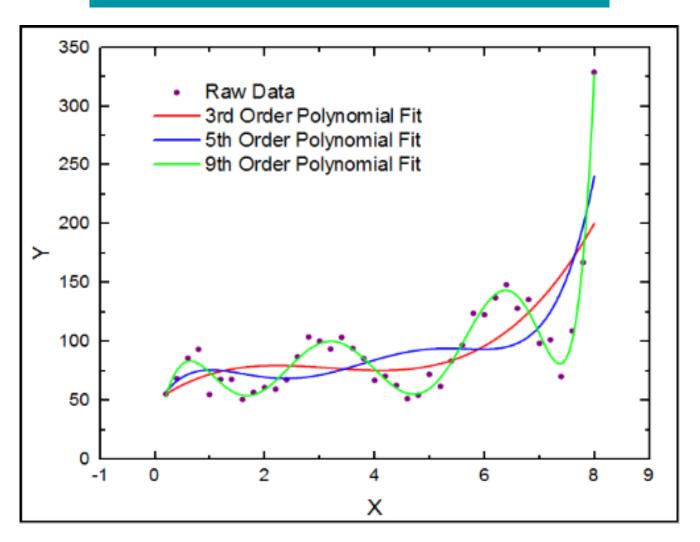


### L'apprentissage machine

$$Y = f_w(X)$$



### Combien de données?





# ...ça dépend

# Idéalement : N échantillons > N paramètres du modèles

- 1. Modèles généralistes pré-entraînés puis spécialisation : Fine tuning / Transfert learning
  - 2. Augmentation de données



#### L'augmentation de données



Duplication de clef

235 références







### **Images de clefs**

### Clef générique

Bruit, flou, rotations, masquage, rouille, etc

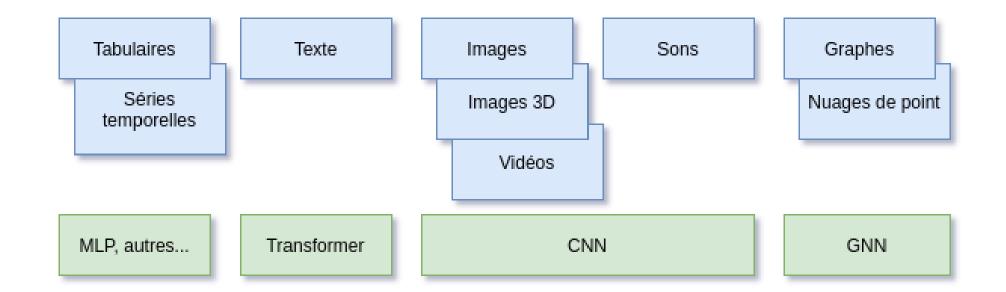
lère suggestion : 95 % de précision 100 % dans les 5 premières < 3 secondes (sur CPU)







## NN & types de données

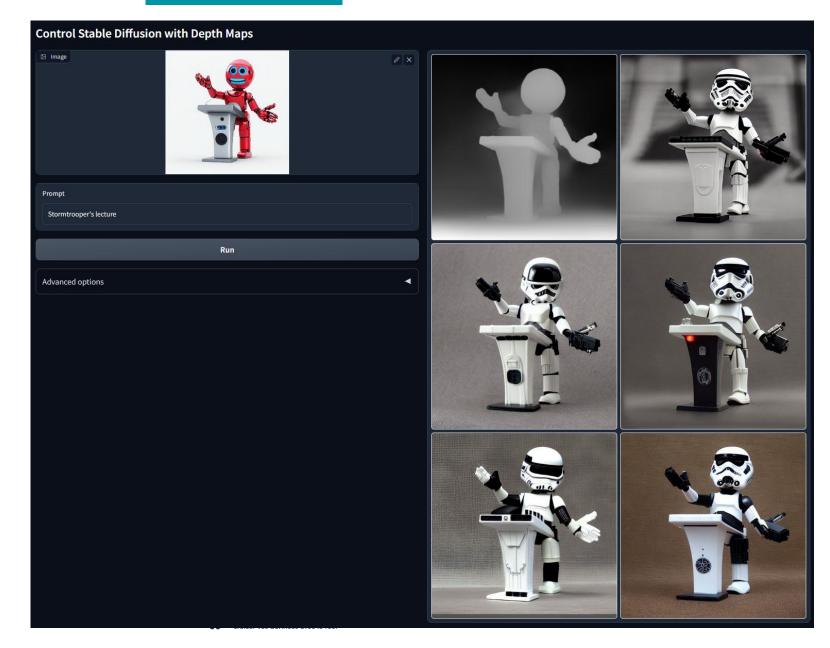






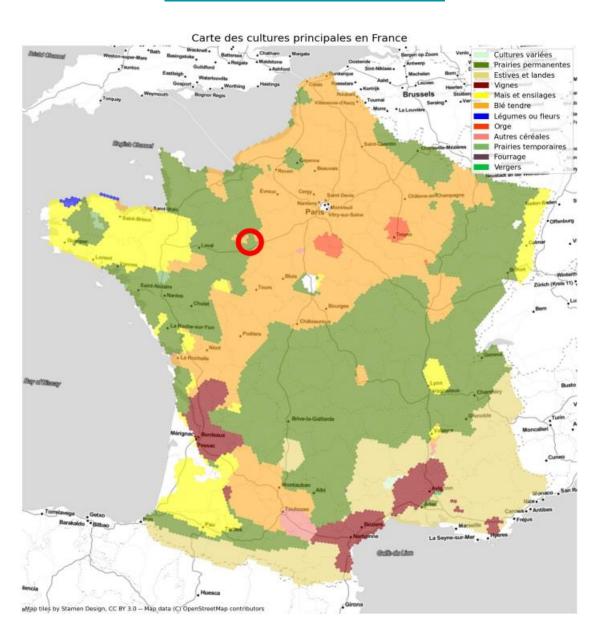
#### Image + Prompt = Image

# Contrôle



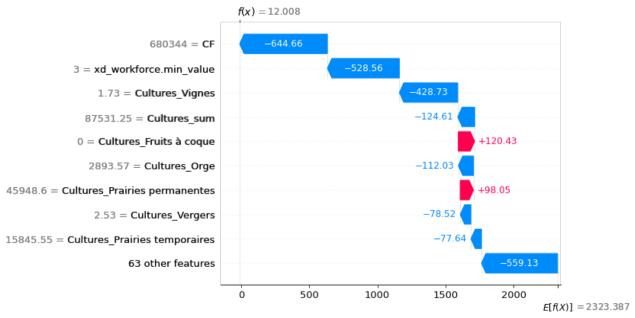
Input Prompt	Question: Explain why this photo is funny? Answer:	Question: Why did the little boy cry? Answer:	Question: What is the hairstyle of the blond called? Answer:	Question: When will the movie be released? Answer:
Completion	The cat is wearing a mask that gives the cat a smile.	Because his scooter broke.	pony tail	On June 27
	(1)	(2)	(3)	(4)
Input Prompt	TorchScale - A Library for Transformers at (Any) Scale	5 + 4  Question: The result is? Answer:	Question: What is the heart	10 12 1 9 3 8 7 6 5 4
c			rate in this picture? Answer:	
Completion	A library that allows transformers to work efficiently and effectively at any scale.	5 + 4 = 9	57 bpm	10:10 on a large clock
	(5)	(6)	(7)	(8)

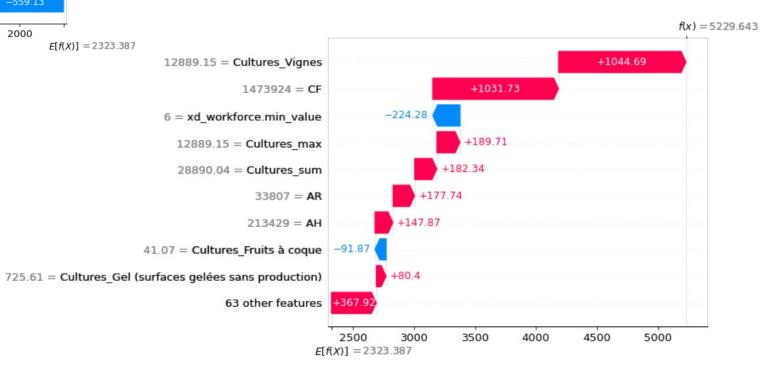
# **Explicabilité**



Cas d'usage

# **Explicabilité**





### Un projet d'IA (réussi)

Maturité « data »

Stratégie / Opérationnel

ROI / Usage





#### **Prédiction**

(IA qui propose à l'humain ou décide à la place)

#### **Analyse**

(Humain qui prend des décisions)

#### **Visualisation**

(Tableaux de bords, Graphiques... etc)

#### Mise à disposition & Traitement

(Base de données commune, architecture pour récupérer les données, granularité, calculs simples)

#### Collecte de données

(ERP, CRM, Machines, données externes ...Etc)

SS DATA

#### Aligner vision stratégique et réponse opérationnelle

Stratégique Demande descendante Bingo! pas adopté / résistance au changement **Opérationnel** Demande Nice to have montante pas utilisé

pas de vision de valeur ajoutée par l'encadrement

### Stratégie : ROI

Indicateurs de succès, mesure & suivi

Cadrage & Data Canvas



### **Opérationnel : l'UX**

Pris en compte dès le cadrage Implication de l'utilisateur final

Ex planification: webgl, filtrage rapide, installation dans atelier





# L'optimisation de planning:

UX

Interface WebGL

**Dans l'atelier** 

Filtrage rapides

Cas d'usage

### Opérationnel : l'explicabilité

Intégration dans l'expérience utilisateur

Fournir assez d'information = confiance dans l'outil



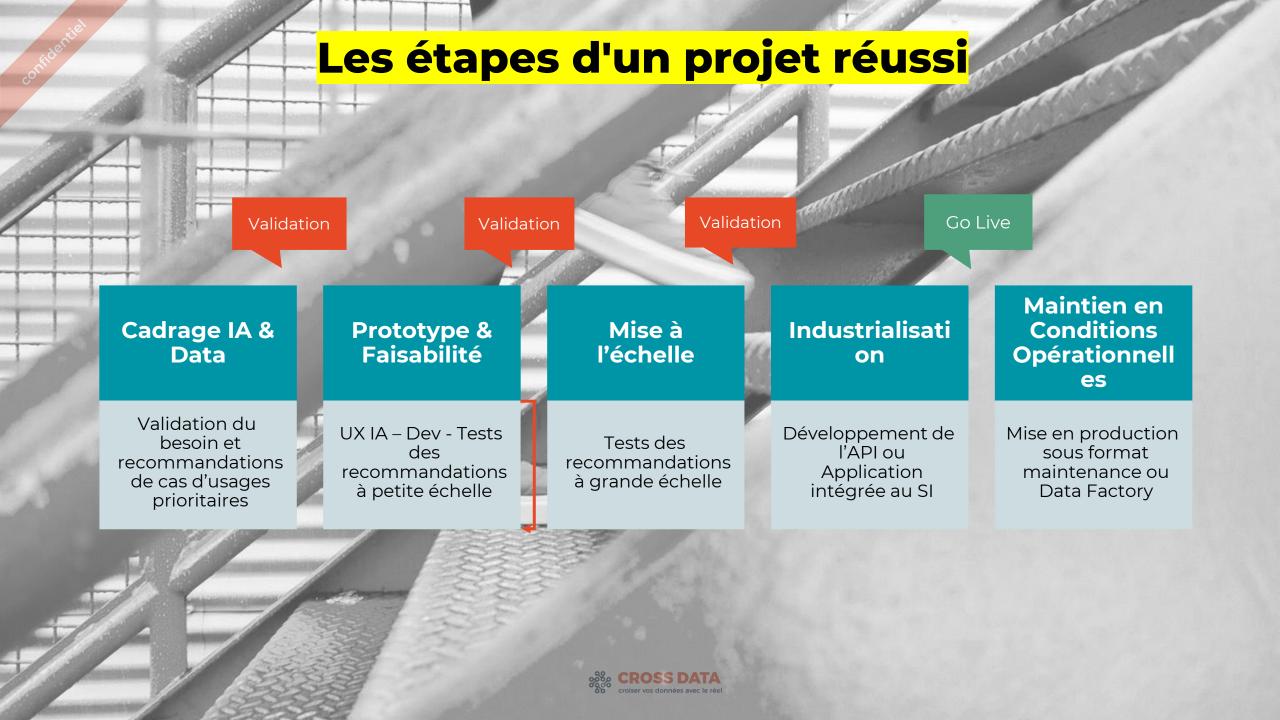
# Maintien en Conditions Opérationnelles

Un peu plus que "ça tourne"...

Contrôle des données

Contrôle des prédictions du modèle





#### Le sujet est vaste & passionnant

(Déploiements des modèles, Human in the loop, Euclidian Neural Networks, etc)

# Disponible toute la journée pour en parler Merci!



Jean-Baptiste Juin ibiuin@crossdata.tech









Angers | Paris | La Roche Sur Yon

02 52 35 08 69 | contact@crossdata.tech | www.crossdata.tech





#### quelques références clients

énergie & environnement

















immobilier





industrie



















retail









transports & supply chain









éducation







éditeurs







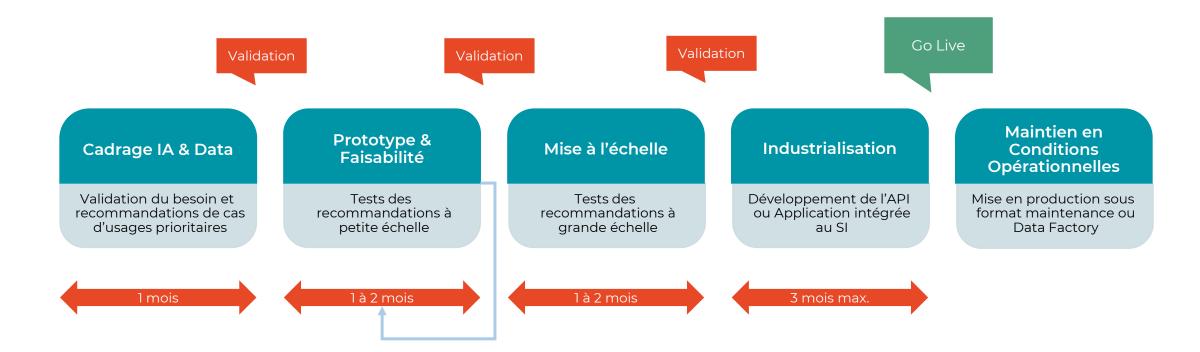
assurances







## un projet réussi chez cross data













énergie



40k pers.

**objectif** identifier un projet rentable

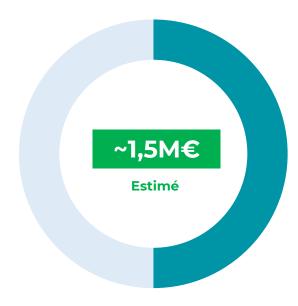
solution diagnostic ia & data & developpement

Dans le cadre du Lab Bretagne, Enedis a identifié trois projets sur lesquels se projeter à l'aide des données.

En 1 semaine, Cross Data intervient pour effectuer un diagnostic des trois projets. Deux projets sont rejetés et un projet est retenu pour partir en développement : l'analyse de fraude avec un ROI attendu entre 200k€ et 3M€.

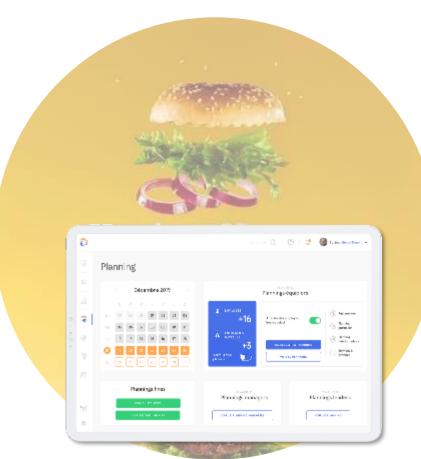
Le projet sera exécuté en 1 mois par Cross Data et son ROI atteint le mois suivant sur la Bretagne.

Depuis le projet a été déployé dans 12 directions régionales.











Planification



40 pers.



Restauration Rapide

## objectif prediction de ventes et de charge

## solution ordonnancement sur mesure & machine learning

RHIS construit une plateforme permettant de gérer les plannings des chaines de restauration rapide Burger King, KFC et Papa John. Un des enjeux forts est de prédire la fréquentation des restaurants afin d'identifier les ressources humaines nécessaires pour servir les clients.

Pour 750 restaurants, Cross Data réussit à prédire leur chiffre d'affaires 10 jours à l'avance par tranches de 15 minutes et par canal de ventes (borne, drive, comptoir) avec un taux de réussite de 97%.

Ce taux de réussite a progressivement dépassé le meilleur manager humain dans les 750 restaurants .











Industrie



200 pers.



Clients B2B

objectif détection de clients à réactiver

solution Machine Learning, Open Data & Facturation

La société Hydrokit vend des solutions techniques à plusieurs milliers d'entreprises spécialisées dans le BTP, l'agriculture et l'environnement en France.

Notre client souhaite identifier

- Des tendances et de corrélations entre types d'entreprises et CA réalisé
- Des tendances et de corrélations entre types de cultures et CA réalisé
- De clients à réengager pour les 50 commerciaux

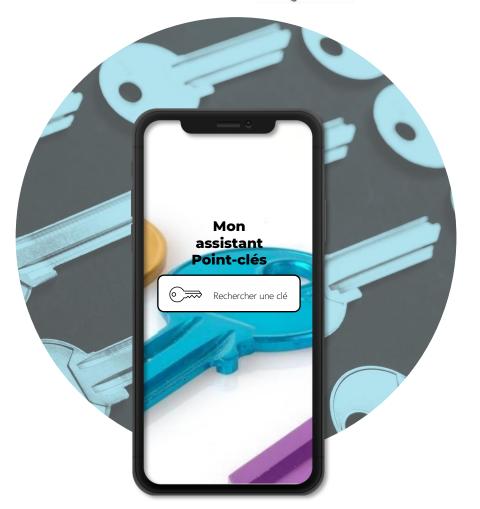
Suite à ce projet, l'entreprise à mis en place un processus de surveillance de ses 16.000 clients et chaque commercial reçoit des alertes sur ses comptes en fonction de leur comportement.













Retail



30 pers.



Clients B2B2C

objectif détection de modèles de clés pour doubles

solution réseaux de neurones sur mesure &

## augmentation de données

L'objectif du projet est de créer une lA capable de reconnaitre les clés à partir d'une photo.

Problème majeur : Un modèle de clé peut avoir 50 combinaisons de recto et versos différents.

Création de la Base d'apprentissage des clés : Prétraitement par Open CV puis augmentation de données pour arriver à un nombre d'images critique par référence (luminosité, étirement, bruit, sans miroirs) : 540k images générées pour

#### 16 références !

Création de deux réseaux de neurones adhoc basé sur TensorFlow et utilisation sous Google Colab pour un résultat supérieur aux attentes du client.





# cas client empovet





Médical





Vétérinaires

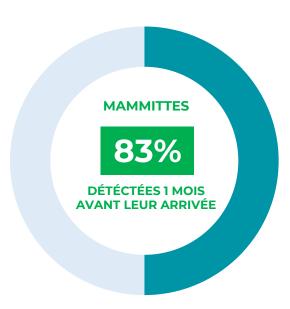
objectif prédiction de maladies chez les bovins

solution Data Science (Stats & ML) sur mesure

Empovet est une startup qui développe des logiciels à destination des vétérinaires ruraux pour les épauler au quotidien dans leur consultation et tâches administratives.

A l'aide des données des contrôles laitiers, des relevés d'exploitation, et des interventions vétérinaires, Cross Data a modélisé des algorithmes sur mesure pour détecter les risques de maladies et notamment de mammites.

Le travail de Cross Data sert aujourd'hui de base aux chercheurs qui peuvent alimenter la plateforme en nouveaux algorithmes.











Industrie



300 pers.



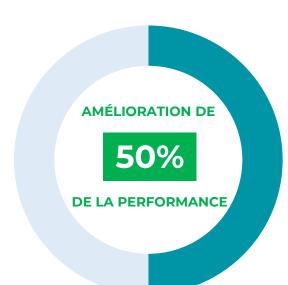
Automobile

## **objectif** optimisation de planification industrielle

#### solution algorithme plne sur-mesure

Dans son usine d'une centaine de machines et de plusieurs milliers de références, notre client connait des problématiques de planification de la production. Avec une performance optimisable, et la sursollicitation des chefs d'ateliers, notre client recherche une solution qu'il pourra internaliser, sans cout de licence prohibitif et permettant un temps de calcul de charge inférieur à 1 minute.

En s'appuyant sur un solveur libre, Cross Data développe une solution sur-mesure qui répond aux besoins du clients en moins de 5,2 secondes (1/10e secondes par semaine) sans changer les méthodes de travail de son client et en complétant des outils connus de ses équipes.











Négoce



250 pers.



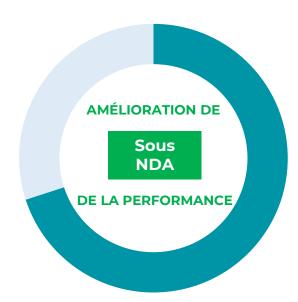
Horticulture

## objectif colisage en 2d avec étages variables

### solution programmation par contraintes

Dans son entrepôt d'expédition de jeunes plants sur des racks, notre client souhaite passer d'un approche picking à une approche de dispatching ce qui nécessite de réaliser les plants des rolls à l'avance et notamment les étages sur les rolls avant de recevoir les plants.

En s'appuyant sur des technologies de programmation par contraintes, Cross Data développe une solution sur-mesure qui répond aux besoins du clients dans le cadre de son changement de process. L'outil estime le nombre de rolls nécessaires à l'avance, définit le positionnement des étagères, puis le positionnement des plans sur les étagères et imprime les rolls pour l'équipe de dispatching.









agriculture



50 pers.



Clients B2B

objectif détection & comptage de racines

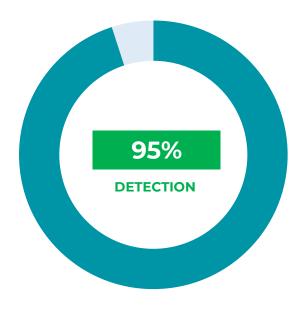
solution réseaux de neurones & analyse d'images

Dans son exploitation de plus de 15 fermes, notre client trie des racines pour ses clients qui les replantent.

Il estime une marge d'erreur de 10% dans ses big bags qu'il corrige en ajoutant 10% du poids pour un risque total de 20% de dépassement.

Cross Data a développé une solution matérielle et logicielle qui lui permet d'assurer un comptage avec moins de 5% d'erreur.

Le comptage est réalisé sur une chaine qui se déplace à 3m/s.











retail





Clients B2C

#### **objectif** performance avec les données commerciales

### solution prévisions & recommandations sur Dataiku

Padd est une société qui dispose de plus 70 Magasins dans toute la France. L'entreprise réalise ses ventes par plusieurs canaux : magasins et site web.

Utilisant déjà une plateforme Dataiku de Data Science, Padd souhaitait vérifier qu'elle avait les bonnes pratiques sur ces sujets.

Accompagné par une Agence de communication, Cross Data a réalisé un inventaire des pratiques data de Padd, listé des <mark>recommandations et des</mark> bonnes pratiques à mettre en œuvre (Design de cohortes). Cette approche a été couplée avec la définition de scénarios de Marketing Automation pour les clients.









Immobilier





Agence digitale

objectif estimation de biens immobiliers

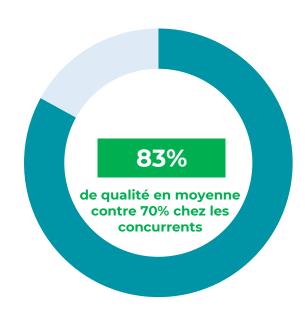
solution MLOps sur mesure dans AWS

Digipi construit la plateforme EstimIA, dans laquelle les agents peuvent estimer en temps réel la valeur d'un bien immobilier

Cross Data produit les algorithmes pour les EstimIA qui permettent à partir des données de l'Open Data et des données de Digipi de réaliser les estimations.

Le défi est d'estimer la valeur d'un bien sans intervention humaine, par l'apprentissage automatique.

Dans le cadre du Covid 19, <mark>le marché immobilier a</mark> évolué rapidement, EstimIA s'y est adapté automatiquement.













Energie





Energéticiens

objectif modélisation des économies d'énergie

solution Clustering & Algos opposables en 6 semaines

Don De chaleur est une application qui utilise les incitations pour faire baisser de 10 à 20% la consommation énergétique des ménages

D2C a contacté CD pour réaliser l'algorithme en 6 semaines, ce que d'autres entreprises du marché ont refusé devant la difficulté et le temps imparti.

Cross Data a réalisé le développement selon une méthode reconnue définie par un spécialiste de l'énergie pour avoir une méthode de calcul d'économies d'énergie de manière opposable.









Formation



50 pers.



Pharmacie

## objectif détection de clients à risque

## **Solution Machine Learning & Salesforce**

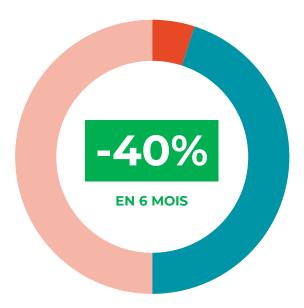
Notre client vend des formations à des pharmacies. Certaines formations sont obligatoires pour les pharmaciens. Le marché de la pharmacie étant limité (environ 20.000 pharmacies), notre client cherche à détecter les pharmacies qui ne vont pas renouveler son offre.

Le défi est d'estimer le risque sans intervention humaine, par l'apprentissage automatique.

Les résultats atteints sont le suivants :

- 5% des entreprises avaient un risque de départ de plus de 50% en année n+1.
- Plus de 10.000 clients analysés
- Baisse de 40% du taux de départ

.







### une expérience éprouvée



Forte de 11 personnes dont **2 docteurs** en mathématiques et astrophysique et **6 data scientists**, Cross Data développe des solutions sur mesure pour les PME, ETI et startups ambitieuses.

Construite sur l'expérience de ses fondateurs au sein du groupe EDF dans l'optimisation des ressources et au CEA dans la détection de galaxies, Cross Data offre aujourd'hui ses compétences en IA aux entreprises ayant un marché B2B.

Les équipes techniques de Cross Data sont formées en recherche opérationnelle dans des **institutions de renommée internationale (ENSTA ParisTech, IMA).** 

Cross Data s'est construite avec un objectif : **Mettre** l'intelligence artificielle au service de la performance des entreprises



### une équipe a la pointe



2 docteurs : 15 ans d'XP en Recherche opérationnelle et IA pour l'industrie et la Recherche

7 Data Scientists



30 % du temps passé en R&D



#### machine learning

Classification, Régression, Clustering : analyse de churn, prévision de CA, algos de similarité

<u>Technologies</u>: Random Forest, Naive Bayes, GAM, Adaboost...etc



#### recherche opérationnelle

Optimisation, ordonnancement, optimisation de tournées/planning

<u>Technologies</u>: PLNE (via branch & bound ou heuristiques telles que algos génétiques, etc.), Programmation par contraintes



#### nlp

Analyse textuelle, Analyse de sentiment

Technologies: Bag of words, TF/IDF, BERT, GPT



#### deep learning / réseaux de neurones

Reconnaissance d'images ou NLP

<u>Technologies</u>: Réseaux convolutionnels (CNN), Transformers, GAN



#### on a envie d'essayer en dehors du laboratoire

Réseaux bayésiens, Reinforcement Learning, Programmation dynamique

