



DATASCIENCE ET INTERACTIONS MÉTIER

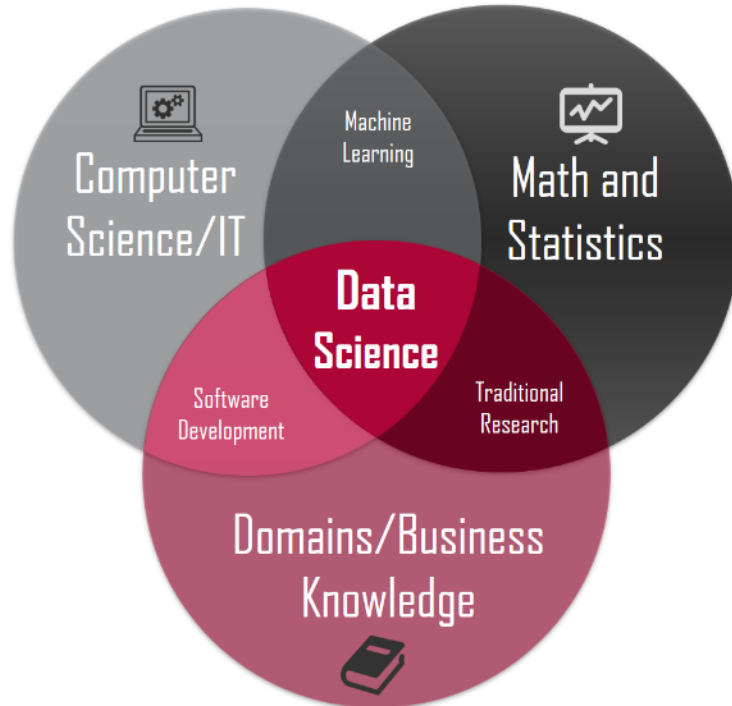




01

INTERACTION MÉTIER

INTÉRACTION MÉTIER



Motivations :

- Dans tous les projets de datascience le datascientiste a **besoin** de la connaissance métier.
 - Forte spécificité des jeux de données et des domaines
 - Forte spécificité des cas d'usages et besoins réels du client
- Le métier a **besoin** d'un minimum de compréhension de la solution développée par le datascientiste
 - Donner de la confiance à une boîte noire est difficile
 - Comprendre les limites est souvent nécessaire à l'utilisation de la solution

Il faut parler la même « langue » et être capable de se comprendre

LES ECHANGES

Avoir des échanges très réguliers avec le métier est **nécessaire** et encore plus en début de projet.

- Dans nos projets : **une fois par semaine ouvrée** (lorsque possible)
 - Fixer des points à date régulière et à l'avance augmente la probabilité que ceux-ci auront bien lieu

3 objectifs lors des interactions découlent des motivations :

- Faire un point d'avancement
 - Assure la bonne progression du projet
 - Maintient un échange avec le métier
 - Définir ce qui doit être fait dans l'imédiat et au long terme
- Obtenir de la connaissance métier
 - Assure une meilleure solution de datascience
 - Assure une solution qui répond au besoin réel
- Vulgariser et donner de la visibilité sur les méthodes et démarches que vous employez
 - Permet une meilleure confiance du métier sur votre travail
 - Augmente l'acceptabilité du travail réalisé

LE POINT D'AVANCEMENT

Nécessaire particulièrement en début de projet

- Souvent plus rapide au fur et à mesure que le projet avance

Dépend de la méthodologie projet

- Nos projets : méthodologie agile avec ses forces et critiques
- Points de durée différentes : suivi, fin de sprint, fin de projet,....

Notes et compte rendu

- Par mail
- Avoir une trace qui peut servir de preuve

Donner de la visibilité sur l'avancement par rapport au planning

- Expliquer ce que vous avez fait par rapport à ce qui avait été prévu de faire la dernière fois
- Expliquer ce que vous allez faire d'ici le prochain point
- Montre au client que vous savez ou vous allez pour avancer sur la réponse à son problème

Arrivez en étant prêt à prendre le lead de la réunion

- Paraître moteur
- Ne prendre le lead que lorsque nécessaire, ne pas l'imposer à tout prix !

ECHANGER AVEC LE MÉTIER

Le métier doit savoir qu'il est écouté et compris.

- Montrez que vous avez compris son problème en le reformulant.
- Montrez que vous comprenez ses données par des visualisations/analyses statistiques.
- Posez des questions si vous n'êtes pas sûr de ce qu'il veut dire.
- Montrez-lui que vous prenez en compte ce qu'il vous apporte dans votre solution.
- Appropriiez-vous et utilisez son vocabulaire dès que possible

Comprenez ce qui apporte le plus de valeur pour le métier et priorisez

- Meilleur modèle ne veut pas nécessairement dire meilleure valeur
- Le métier peut ne pas savoir ce qui lui apportera le plus de valeur
 - Parce qu'il ne connaît pas l'IA et ce que cela peut lui apporter
 - Parce qu'il croit trop à l'IA
 - Parce qu'il n'y a pas pensé

Mettez-vous à sa place, essayez de le questionner pour mieux comprendre.

- *Comment il fonctionne habituellement ? Quels sont ses "pains points" ?*
- *Comment il résout le problème actuellement ? Quel est l'existant ?*

VULGARISATION DES MODELES ET METHODES

- **Expliquer le rôle et le fonctionnement de l'IA en général peut être utile**
 - Eviter les idées reçues
 - L'IA remplace rarement un job, mais permet de mieux le faire
 - Il est préférable que ce soit le plus tôt possible dans le projet
- **Soyez honnête et transparent :**
 - si le modèle ne fonctionne pas comme espéré, dites-le, et expliquer les causes possibles et les actions envisagées pour essayer de corriger ça
 - Si le modèle a peu de chance de fonctionner au vu de facteurs comme le manque de donnée, ...
 - Ne pas tricher sur l'évaluation, cela se verra en production
 - Convaincre le métier de la rigueur de votre évaluation est un plus !
- **Définissez les termes**
 - Les termes techniques
 - Certains termes peuvent être connus de tous mais avec des sens différents
- **Décidez le niveau d'explications fournies**
 - Vous ne pouvez pas tout expliquer
 - Fixer le niveau suffisant pour comprendre le principe de ce que vous faites
 - Tous les détails ne sont pas nécessaires, mais peuvent au contraire perdre votre interlocuteur
 - Dépend de la personnalité et background technique de la personne en face



02

**CANVAS MACHINE LEARNING
- LOUIS DORARD**

The Machine Learning Canvas (v0.4)

Designed for:

Designed by:

Date:

Iteration:



Domain Integration

Predictive Engine



03

**DIAGNOSTIQUER UN CAS
D'USAGE AVEC LE MÉTIER**

EVALUER UN CAS D'USAGE

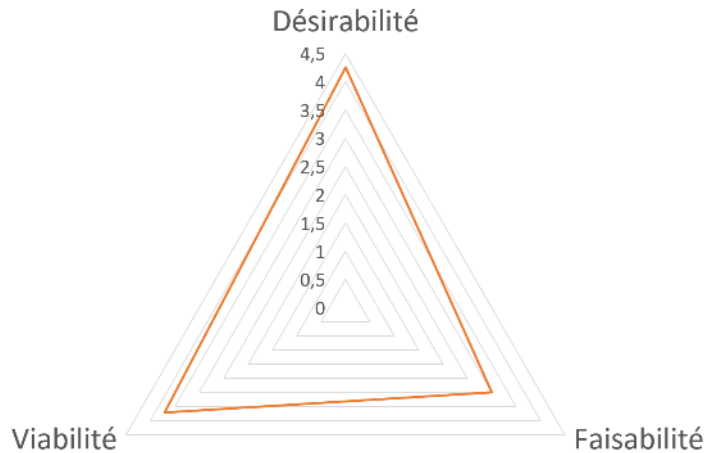
- Comment décider si un cas d'usage est intéressant, et mérite un projet ?
- Comment sélectionner le meilleur cas d'usage ?

3 critères :

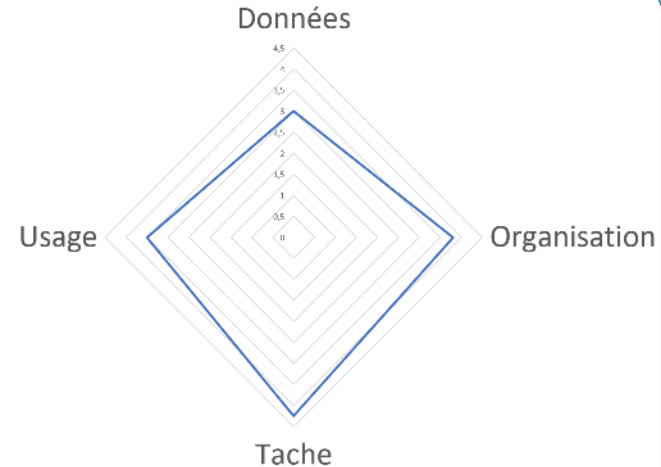
- **Faisabilité**
 - Est-ce réalisable par rapport aux contraintes du projets ?
 - Contraintes données, cas d'usage, temps, performances attendues, ...
- **Viabilité**
 - Quelle est la valeur apportée ?
 - Quel est le cout de la solution ?
 - Sa durée de vie ?
- **Désirabilité**
 - A quel point la solution va être désirée par son utilisateur ?
 - Caractère pénible/intéressant de la tâche
 - Quelles sont les impacts sur la vie de l'utilisateur ?
 - Comble un besoin ressenti par l'utilisateur ?

NOTRE QUESTIONNAIRE

Diagnostic à partir de la feuille



- Viabilité et Désirabilité marquées, l'intérêt du cas d'usage est établi et semble bien perçue par les métiers.
- On peut observer toutefois un point de vigilance sur la faisabilité du projet qui est liée aux données et à la mise à disposition d'un expert métier.



- La nature de la tâche et le contexte organisationnel ont un impact favorable sur le projet.
- Point de vigilance sur la facilité d'accès aux données, la quantité de cas anormaux et en l'annotation de ces derniers. Notamment à la lumière du caractère chronophage de l'anonymisation.

**Ce diagnostic se base sur la grille d'analyse des Uses Cases développée par SCALIAN pour valider les projets IA.*

Les actions conséquentes

PLAND'ACTION

- Afin de réduire les risques de non-faisabilité il nous faut **planifier des temps d'échanges réguliers entre les experts métiers** capable de détecter une anomalie et l'équipe data science chargée d'entraîner les modèles de détection.
- La **mise à disposition de l'ensemble des données nécessaires** à la détection d'une fraude est le critère le plus prioritaire. Ces données doivent suffire à un expert sans connaissance extérieure de pouvoir détecter une fraude.



04

**BONUS : ANALYSER UN CAS
D'USAGE IA EXISTANT**

EVALUER UN CAS D'USAGE IA EXISTANT

Les circonstances :

- Pour estimer l'intérêt du cas d'usage
 - Pour s'en inspirer ou le proposer dans un domaine similaire
- Pour estimer l'intérêt de la solution technique déployé
 - Pour évaluer son intérêt dans un domaine ou tâche similaire
 - Pour voir ce qui a marché, ses limites et à quoi faire attention
 - Pour décider l'intérêt de solutions sur étagères

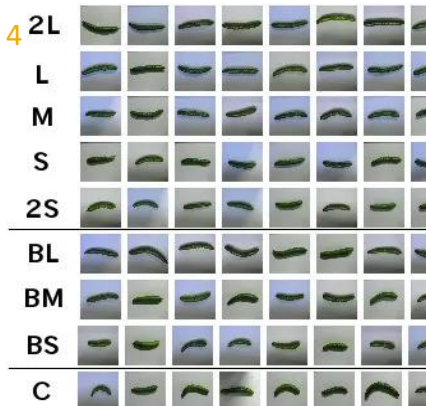
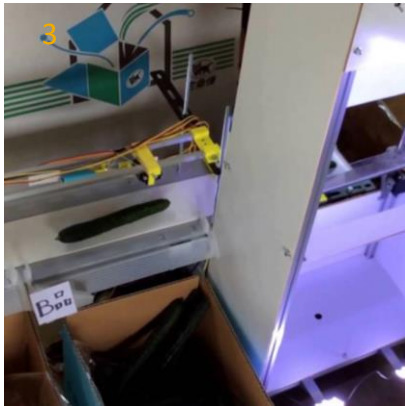
- **Souvent sans accès au métier du cas d'usage**

- **Souvent sans accès aux données**

- **Souvent peu d'information sur le contexte exact**

TRIER LES CONCOMBRES

L'INTELLIGENCE ARTIFICIELLE AU
SERVICE D'UNE FERME JAPONAISE



Contexte

Une fois cueillis, les concombres de cette ferme doivent être triés en 9 catégories selon leur forme et leur couleur. Les circuits de distribution est les prix ne seront pas les mêmes pour ces 9 types de concombre.

Problème

Le tri des concombres, jusqu'à 4000 par jour, sollicite une employée qualifiée à temps plein pendant le pic de production.

Solution

Le fils de la famille a implémenté une solution à base de computer vision (2) qui détecte le type de concombre. Assortie d'un convoyeur automatique (3) capable de trier les concombres selon les catégories identifiées.

Statut

POC artisanal. Le projet des défauts qui le rendent inutile en pratique, mais illustre un cas d'usage pertinent qui existe en production dans d'autres contextes.

Attention !!



L'histoire est plaisante et a beaucoup été relayée. Néanmoins, la solution est moins utile qu'elle en a l'air. Le système de tri physique n'est pas très performant, et la précision du modèle encore insuffisante. Cette histoire sert néanmoins à illustrer un cas d'usage parfaitement traité par des solutions plus abouties industriellement.

A retenir

La démocratisation du Deep Learning : le mouvement opensource avec Tensorflow rend le DeepLearning plus accessible que jamais. Le valeur de l'expertise technique : la stratégie employée (transfer learning basé sur MNIST) n'est probablement pas très adaptée et explique la performance décevante (70%).

Pourquoi ça aurait pu marcher ?

Les tâches de tris sont de bons sujets pour l'intelligence artificielles, d'autant quand les différences sont visibles (comme c'est le cas pour les concombres).

La principale force du projet, c'est la stratégie d'acquisition de données labélisées (photos de concombres déjà triés) qui a permis d'entraîner le modèle sur plus de 7000 images. Probablement encore insuffisant, mais impressionnant dans un contexte de POC DIY. L'intégration de la brique intelligente de tri aurait permis d'automatiser de bout en bout la tâche dévolue à la mère.

Les risques ?

Tâche complexe, pas forcément de features adaptées. Impact des erreurs sur la suite de la chaîne d'approvisionnement ?

Make or buy ?



MAKE

Spécificité de la tâche

Jeu de données existant

Méthode connue et
éprouvées sur d'autre
domaines

BUY

Solutions existantes ?

Machine de tri

Sources

<https://www.newyorker.com/tech/annals-of-technology/diy-artificial-intelligence-comes-to-a-japanese-family-farm>

- *Appropriation par google du cas d'usage*

Cas similaires

L'usine automobile



Contexte

Dans les usines les défauts de fabrication engendrent d'importants surcoûts. Être capable de détecter uniformément une pièce mal désignée, non conforme le plus rapidement possible est un enjeu majeur. Les contrôles sont effectués manuellement.

Problème

Face à la grande volumétrie de pièces usinées, il est complexe de maîtriser la qualité des produits en cours de fabrication et d'uniformiser les notes données par les différents collaborateurs.

Solution

La visions par ordinateur liée à des modèles de clustering permettent de détecter avec une très grande finesse les défauts de qualité. Ainsi le modèle fournit une note de non-qualité pour chaque pièce. Si un type de pièce est trop souvent mal notée, alors les designers la retravaillent?

Statut

Le produit est en production dans plusieurs usines du groupe FAURECIA. En 2020 ils espèrent équiper 15% de leurs usines. Cette application permet de réduire de 50 % les défauts sur les pièces contrôlées (par échange avec les designers), ainsi que des gains de temps et des réductions de coûts importantes.

L'œil critique

Attention !!

Le coût de la mise en place d'une telle solution n'est pas forcément intéressante pour toutes les industries. Comme tout outil d'intelligence artificielle il doit être accepté et compris par les métiers. Un tel outil pour être exploité au maximum doit modifier une grande partie de la chaîne de production.

A retenir

Les grandes industries font de plus en plus confiance à l'IA et parviennent à le mettre en production à très grande échelle.

Pourquoi ça fonctionne?

La vision par ordinateur est une des applications les plus maîtrisées aujourd'hui en machine learning. Les modèles développés par la recherche quelques années auparavant ont été largement testés par les industries. Il existe de nombreuses ressources open source permettant d'obtenir des modèles performants.

Risques

Les défauts d'une pièce peuvent potentiellement être très variés ce qui peut être complexe à capter.

De plus l'explicabilité de tels modèles n'est pas toujours évidente. Si une pièce est détectée comme défectueuse mais que le collaborateur n'identifie pas la raison, la pertinence et la confiance d'un tel outil peut être rapidement remis en cause. Un tel outil doit être accepté par les personnes qui l'utilisent.

Sources

<https://www.larevuedudigital.com/faurecia-teste-lintelligence-artificielle-watson-jusqua-present-tout-va-bien/>

<https://www.linkedin.com/jobs/view/developer-internship-at-faurecia-1755265869/?originalSubdomain=fr>

<https://www.usinenouvelle.com/article/faurecia-affine-ses-inspections-grace-a-l-ia.N877315>

<https://www.manufacturing.fr/faurecia-intelligence-artificielle-et-humaine-collaborent/>

Cas similaires

Intel : <https://www.edge-ai-vision.com/2018/10/vision-data-and-ai-connecting-the-factory-of-the-future/>

Scortex.io : <https://scortex.io/>

Scortex : <https://www.meetup.com/fr-FR/Bordeaux-Machine-Learning-Meetup/events/271160321/>

FieldBox

<https://devmesh.intel.com/projects/manufacturing-quality-control-using-deep-learning-computer-vision>

Make or buy ?



MAKE

Anomalie complexe à
identifier

Objets spécifiques

BUY

Objets génériques
susceptibles d'être
couverts par les solutions
sur étagères