



Présentation Probayes

- Recrutement

LPSC, Mardi 03/05/2022

Ensemble, construisons vos solutions d'Intelligence Artificielle

Intervenants

Vincent Bonnivard

DATA SCIENTIST / CHEF DE PROJET

Probayes – depuis 2016

Parcours universitaire :

- L1-L2 Physique/Maths - **Université de Savoie**
- L3 Physique - **Université de Grenoble**
- Master Recherche Physique Subatomique et Astroparticules - **Université de Grenoble**
- Doctorat en Physique des Astroparticules – **LPSC, Université de Grenoble**



Alexandre Ghelfi

DATA SCIENTIST / CHEF DE PROJET

Probayes – depuis 2016

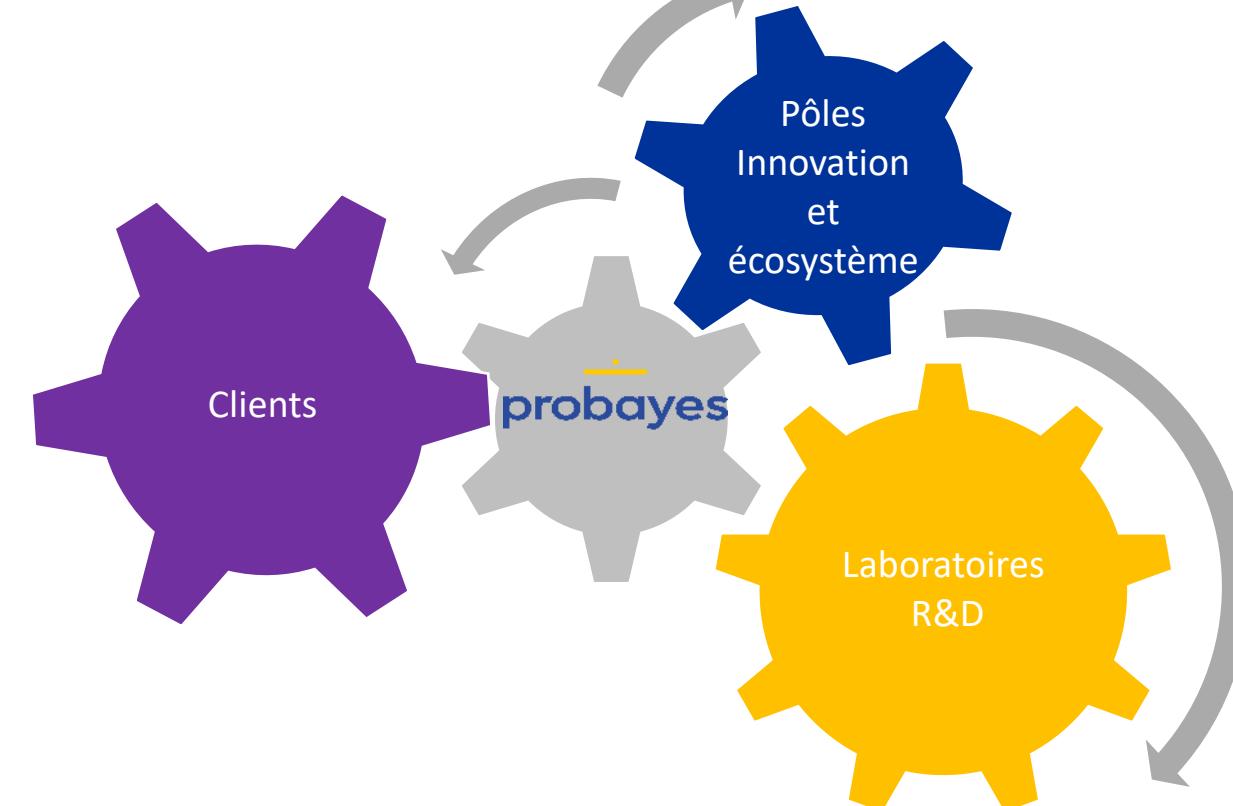
Parcours universitaire :

- L1-L3 Physique - **Université de Grenoble**
- Master Recherche Physique Subatomique et Astroparticules - **Université de Grenoble**
- Doctorat en Physique des Astroparticules – **LPSC, Université de Grenoble**



Sommaire

1. Présentation de Probayes
 1. Organisation
 2. Quelques projets d'I.A.
2. La réalisation d'un projet I.A.
3. Offre d'emploi
4. Questions / Réponses



- Crée en 2003 - filiale Groupe La Poste depuis 2016
- Spécialiste en Intelligence Artificielle
- Essaimage de l'INRIA & CNRS
- Au cœur d'un écosystème pointu de Recherche et Développement
- Des liens forts avec le monde académique :
 - Projets collaboratifs / Thèses CIFRE
 - Conseillers scientifiques
- 62 personnes
- Siège à Grenoble + 1 équipe à Paris



4 pôles de compétences

- Vision par ordinateur
- Traitement automatique des langues (NLP)
- Apprentissage automatique « classique »
- Recherche opérationnelle

4 équipes techniques, dirigées par des directeurs de projet

- 2 équipes spécialisées : **Vision par ordinateur, Recherche Opérationnelle**
- 2 équipes généralistes

+ Fonctions support :

- Admin / RH
- IT
- Commerciaux
- Communication / Marketing

2 types principaux de profils au sein des équipes techniques :

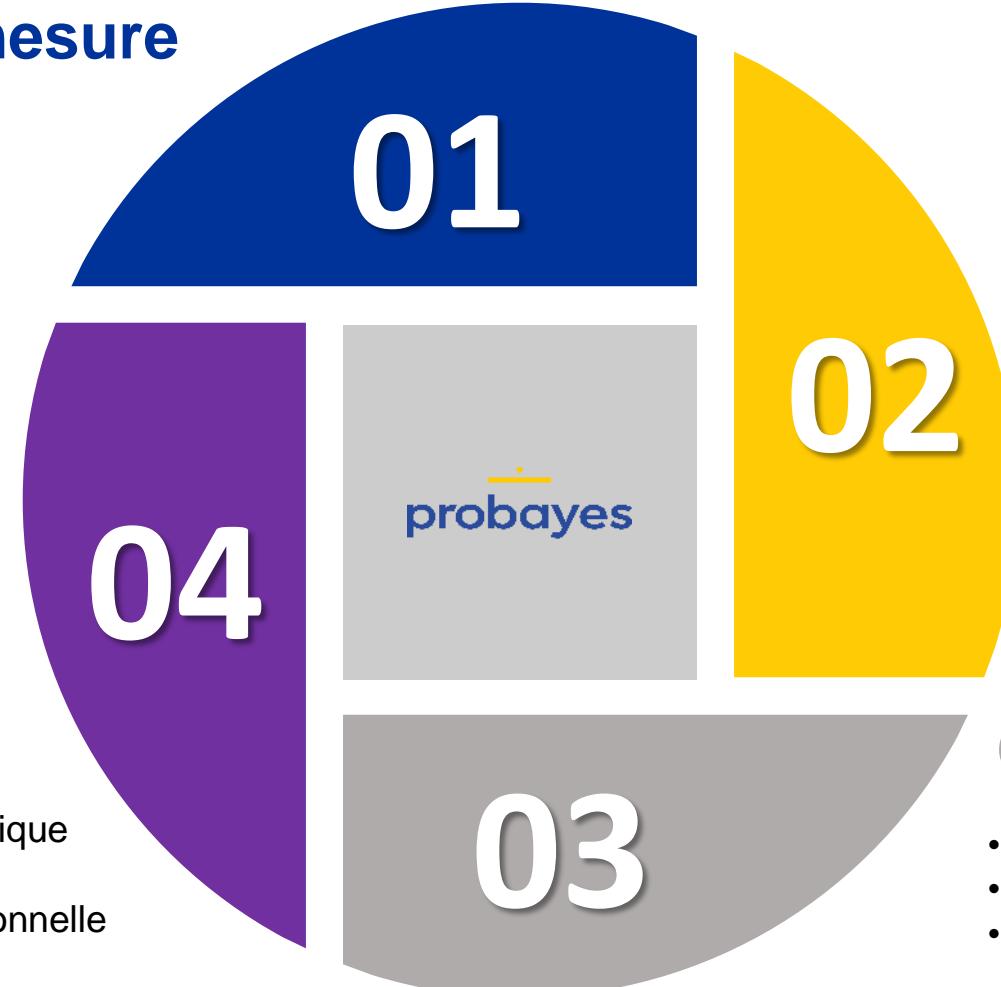
1. Profil "**Data Scientist**" : Ingénieur ou Docteur spécialisé en analyse de données et Machine Learning
2. Profil "**Développeur**" : Ingénieur en développement informatique, avec un savoir-faire en déploiement de solution I.A.

Actuellement, ≈15% des profils techniques sont "Développeurs"



Solutions sur mesure

- Automobile
- Banque et Assurance
- Défense
- Industrie
- Logistique et Distribution
- Santé
- Etc.



Formations

- Machine Learning
- Traitement Automatique des Langues
- Recherche Opérationnelle
- Python
- Etc.

Produits

- BatiSense
- Chatbot
- FraudIA

Conseil

- Définir les cas d'usage
- Identifier les données
- Comprendre le process de mise en application d'une application d'IA
- Acculturer à l'IA

Quelques projets d'I.A.



FraudIA™

Des transactions bancaires en toute sécurité

ENJEUX

- Réduire les pertes financières liées aux fraudes tout en minimisant la gêne client
- Capacité à traiter une forte volumétrie de données quotidienne de transactions cartes bancaires
- Capacité à disposer d'une rapidité de traitement en réponse à une demande d'autorisation

DÉMARCHE

- Un outil de scoring en temps réel des transactions par cartes bancaires pour apporter une réponse d'autorisation
- Une approche d'intelligence artificielle apprenante qui s'adapte donc aux évolutions des pratiques et comportements des fraudeurs
- Une architecture big data capable de passer à l'échelle



Jusqu'à 35%

De rappel de fraudes non détectées par les autres outils



7,5 millions

De transactions par jour pour 4 établissements bancaires



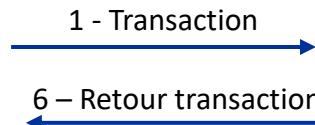
50ms

De temps de réponse max.

RÉSULTATS

- Gain de 10 à 20% (en montant)
- 7 millions de transactions/jour
- Temps de réponse de 50 millisecondes
- Fort niveau de sécurité informatique
- Déployé chez la Banque Postale et la Société Générale (20% des transactions en France)

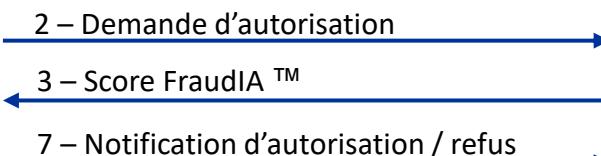
Achat en ligne



Serveur d'autorisation émetteur



Progiciel de fraud management - ISoft



FraudIA™



SOCIETE GENERALE

Computer vision – Fusion de capteurs

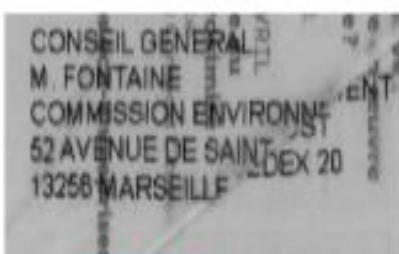
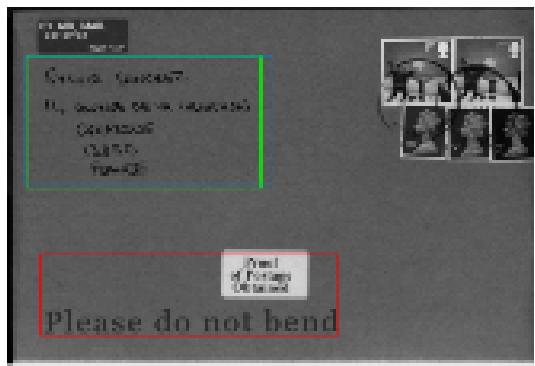
Reconnaitre des adresses

ENJEUX

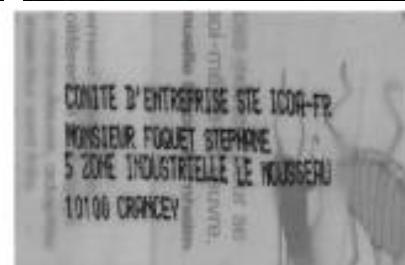
- Améliorer la reconnaissance automatique des adresses sur les plis Petit Format (PF) et Grand Format (GF)
- Réduire les coûts du vidéo-codage manuel utilisé pour les cas d'échec de la reconnaissance automatique

 = OCR actuel

 = Probayes



PREDICTION:
CONSEIL GENERAL
M FONTAINE
COMMISSION ENVIRONNEMENT
52 AVENUE DE SAINT INDEX 20
13256 MARSEILLE



PREDICTION:
CONITE D'ENTREPRISE STE ICOAA FR
MONSIEUR FOQUET STEPHANE
5 ZONE INDUSTRIELLE LE MOUSSEAU
10100 CRANCEY

DÉMARCHE

- Utilisation de méthode réseau profond récurrent, état de l'art OCR pour la lecture de panneaux de rues
- Portage de l'état de l'art dans le cadre d'une procédure d'OCR générique capable à partir d'une image d'un pli petit format de produire l'adresse complète associée.
- Création d'une base d'apprentissage annotée manuellement

RÉSULTATS

- Déployé dans 90% des plateformes industrielles courrier
- Diminution de plus de 40% des requêtes de vidécodage
- Augmentation taux de mécanisation de 10 pts
- Un ROI annuel à plusieurs M€



Traitement Automatique des Langues

Générer automatiquement des imputations comptables

ENJEUX

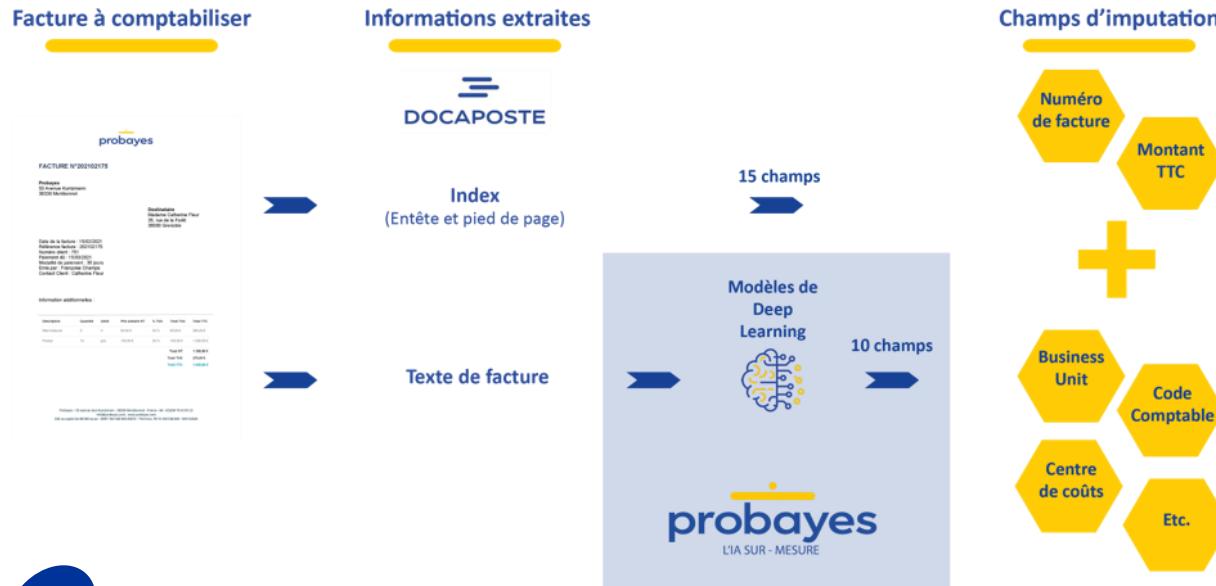
- Les comptables du Groupe La Poste traitent manuellement plus de 600 000 factures fournisseur par an
- Ce travail représente une importante charge pour les comptables
- Le Groupe souhaite rendre automatique la génération des champs d'imputation comptable
- La solution doit fonctionner pour les 3 ERPs du Groupe

DÉMARCHE

- Entraînement de modèles Deep Learning à partir des couches texte des factures, extraites via OCR par Docaposte/Seres, avec une historique de plus de 300 000 factures
- 10 champs de type "classification" prédits : Compte de Charge, Code TVA, Centre de coût... Ce sont des informations déduites de la facture (non-présentes dans le document)
- Les informations présentes sur la facture (Montant, date...) sont extraites par Docaposte/Seres

RÉSULTATS

- Plus de 80% des factures peuvent être traitées automatiquement par les modèles
- Les 20% restants sont identifiées par les modèles, et sont traitées manuellement par les comptables
- La solution est en production dans le DataLake du Groupe La Poste
- Volonté de proposer la solution à l'extérieur du Groupe



Machine Learning

Optimiser le procédé de production

ENJEUX

- Déetecter les conditions non-optimales de production d'anodes
- Afficher en temps réel les facteurs d'influence
- Proposer des recommandations d'actions pour retourner vers l'optimum, en temps réel



DÉMARCHE

- Calculs d'indicateurs résumant le comportement du procédé (40 séries temporelles)
- Modèle de classification pour détection de périodes de sous-densité
- Application de **modules d'explicabilité** : obtention des facteurs d'influence en temps réel
- Développement d'un **module de recommandation**, basé sur un modèle de régression et employant la liste des paramètres actionnables
- Développement d'une IHM dédiée

PROJET

- Un de cas d'usage d'un projet européen H2020
- POC, puis démonstrateur, puis industrialisation
- La solution est déployée dans l'usine d'Aluminium Dunkerque

RÉSULTATS

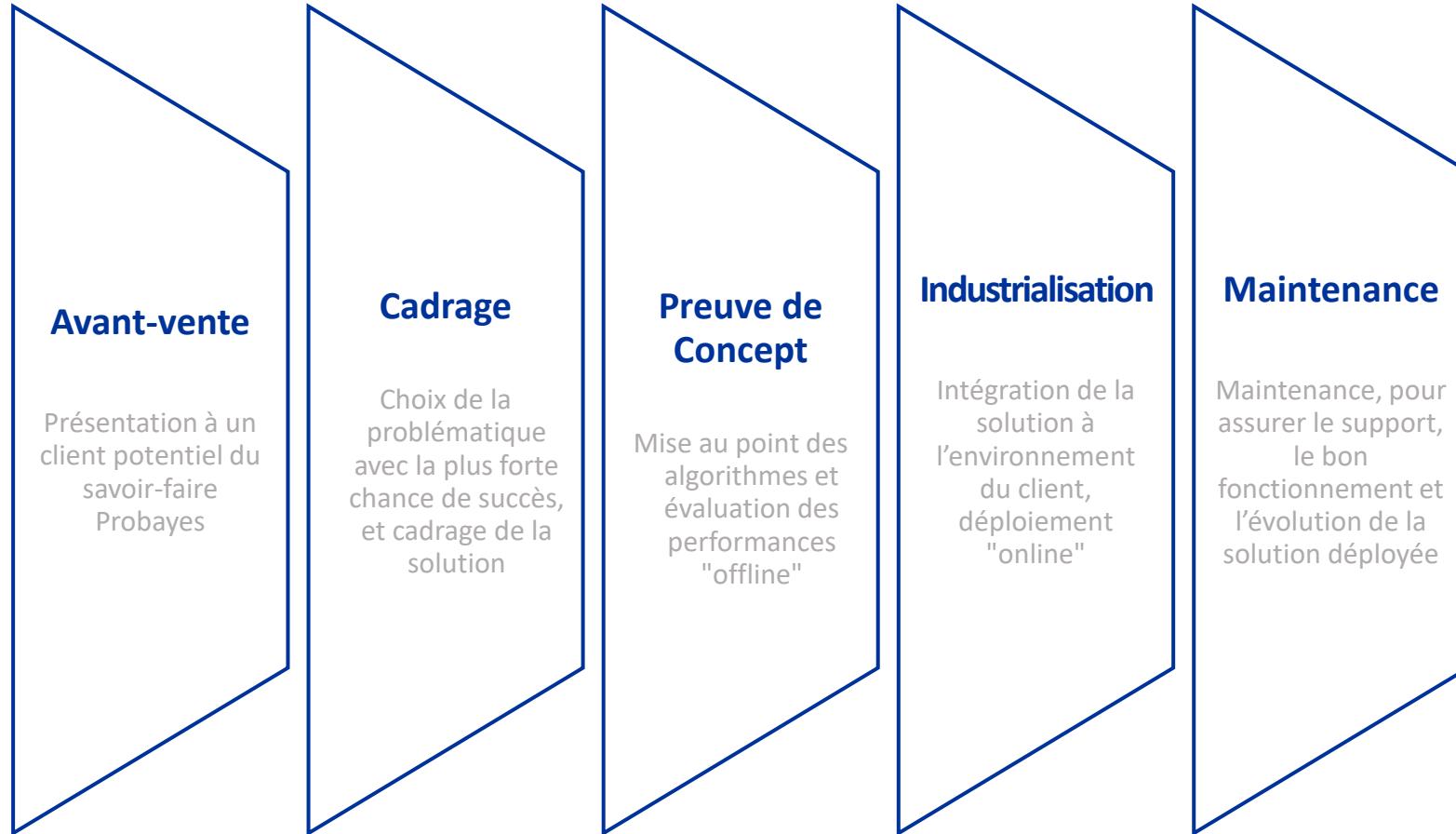
- Solution en production depuis Septembre 2019
- Détection de 70% des périodes de sous-densité
- En temps réel, l'IHM affiche les 5 principaux facteurs d'influence sur la sous-densité, par période détectée
- Une recommandation d'action (3 paramètres à modifier) est proposée
- Gains théoriques estimés à plusieurs centaines de k€ par an

La réalisation d'un projet I.A.



La réalisation d'un projet I.A.

Les phases du projet



La réalisation d'un projet I.A.

Avant-vente

Les commerciaux **contactent leur réseau ou sont contactés**. Un rendez-vous est organisé :

- Déplacement chez le client, **commercial accompagné d'un Data Scientist expérimenté**
- Présentation de Probayes et des projets pouvant potentiellement intéresser le client

Suite à ce rendez-vous, le client va potentiellement revenir vers nous avec des **cas d'usage** à proposer à Probayes : **on démarre alors la phase de cadrage**



La réalisation d'un projet I.A.

Cadrage

Le cadrage consiste à définir précisément le périmètre :

- Le périmètre fonctionnel de la solution : **quels résultats attendus ? Quels utilisateurs ?**
- Le périmètre de **données** : **en quoi consistent les données ? Quelle profondeur d'historique disponible ?**

Une fois le périmètre défini, Probayes va **réaliser une proposition commerciale** :

- Un chef de projet va estimer le **nombre de jours** (= la charge), le **délai** et les **profils** nécessaires pour réaliser la **Preuve de Concept**
- Le commercial va proposer une enveloppe budgétaire associée

Remarque : le projet est réalisé en général **au forfait**

Exemple : Probayes chiffre 50 Jours de travail pour le projet; mais le projet peut se réaliser effectivement en 30 comme en 70 jours de travail !



La réalisation d'un projet I.A.

Preuve de Concept

La Preuve de Concept (Proof Of Concept, POC) est l'étape clé du projet
 --> c'est là que les **algorithmes d'I.A. sont développés et évalués**

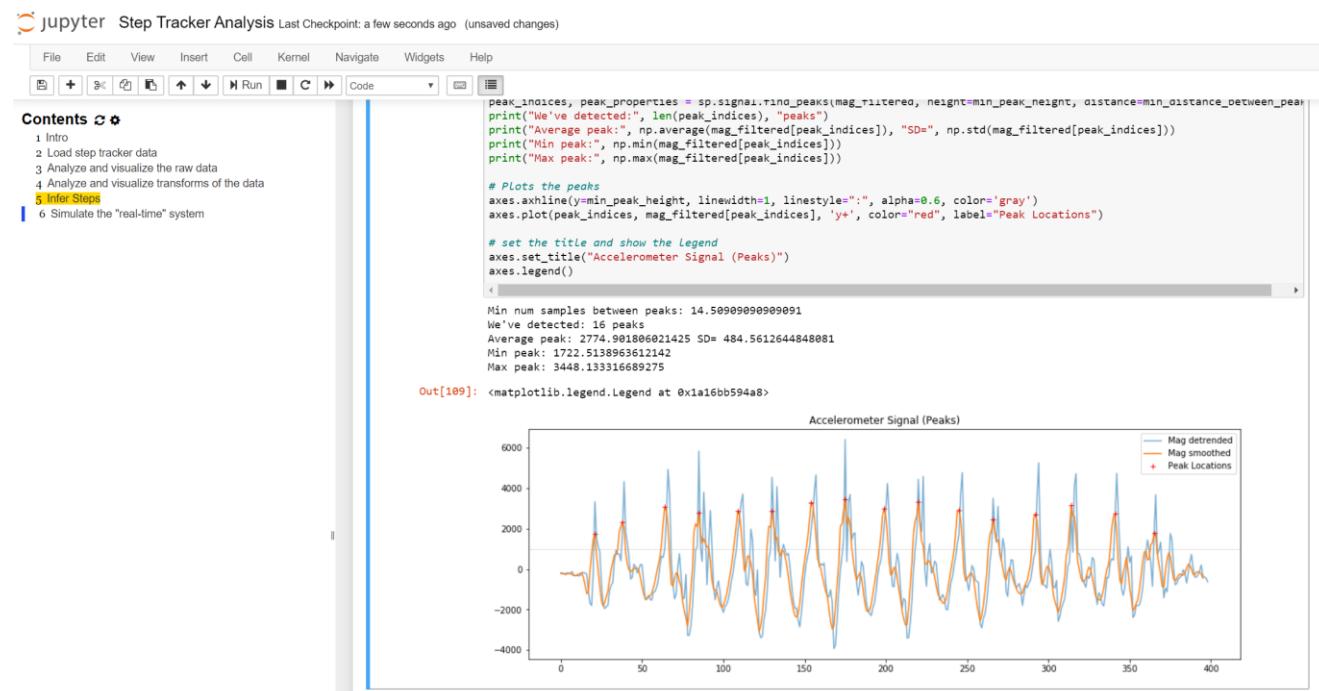
En pratique :

1. Récupération et analyse des données :

- Visualisation des données
- Analyse statistique descriptive
- Détection des valeurs manquantes / aberrantes
- ...

Outils employés :

- Python
- Bibliothèques : matplotlib, pandas, numpy, ...
- Jupyter-notebook
- Gitlab : **code versioning**



The screenshot shows a Jupyter Notebook interface with the following details:

- Title:** Step Tracker Analysis
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Navigate, Widgets, Help
- Code Cell:**

```

peak_indices, peak_properties = sp.signal.find_peaks(mag_detrended, height=min_peak_height, distance=min_distance_between_peaks)
print("We've detected:", len(peak_indices), "peaks")
print("Average peak:", np.average(mag_filtered[peak_indices]), "SD=", np.std(mag_filtered[peak_indices]))
print("Min peak:", np.min(mag_filtered[peak_indices]))
print("Max peak:", np.max(mag_filtered[peak_indices]))

# Plots the peaks
axes.axhline(y=min_peak_height, linewidth=1, linestyle=":", alpha=0.6, color='gray')
axes.plot(peak_indices, mag_filtered[peak_indices], 'y+', color='red', label="Peak Locations")

# set the title and show the legend
axes.set_title("Accelerometer Signal (Peaks)")
axes.legend()

```
- Output Cell:**

```

Out[109]: <matplotlib.legend.Legend at 0x1a16bb594a>

Accelerometer Signal (Peaks)

Mag detrended
Mag smoothed
Peak Locations

```

A plot titled "Accelerometer Signal (Peaks)" is displayed below the code. The plot shows three data series: "Mag detrended" (blue line), "Mag smoothed" (orange line), and "Peak Locations" (red '+' markers). The x-axis ranges from 0 to 400, and the y-axis ranges from -4000 to 6000.

La réalisation d'un projet I.A.

Preuve de Concept

2. Ateliers avec les experts métier : **s'assurer de la compréhension fine du cas d'usage et des données**

- Présentation de l'analyse des données
- Questions/Réponses autour du cas d'usage et des données

--> cette étape est primordiale ! Impossible de réaliser un projet d'I.A. sans comprendre finement les détails du sujet

Lors de ces ateliers :

- Les experts métier du client progressent en I.A. et Data Science
- Les Data Scientist Probayes progressent dans le domaine du client



La réalisation d'un projet I.A.

Preuve de Concept

3. Préparation des données

Avant de réaliser l'apprentissage des modèles, il est nécessaire de **préparer les données** :

- La préparation dépend fortement de la nature de la donnée en entrée : séries temporelles, images, documents...
- Suivant la nature, il est nécessaire de **nettoyer les données, calculer des features, appliquer des prétraitements** (normalisation, scaling, ...), **filtrer des cas hors périmètre** etc...

C'est une étape chronophage, mais nécessaire.



La réalisation d'un projet I.A.

Preuve de Concept

4. Apprentissage et évaluation des modèles : l'étape clé de la Preuve de Concept

- Séparation des données en jeu d'apprentissage et jeu(x) d'évaluation
- Définition des métriques de performance
- **Choix du ou des modèles de Machine/Deep Learning à employer**
(avec revue préalable de la littérature, éventuellement)
- **Apprentissage des modèles** - optimisation des hyperparamètres
- Evaluation des performances sur les jeux d'évaluation

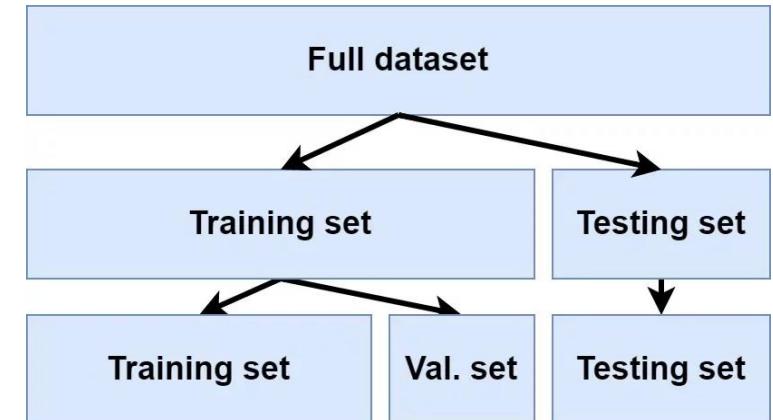
Outils et bibliothèques (Python)

Machine Learning :

- Scikit-Learn
- XGBoost
- ...

Deep Learning :

- Tensorflow
- PyTorch
- HuggingFace
- ...



La réalisation d'un projet I.A.

Preuve de Concept

5. Recette des performances par le client

- Le client évalue les performances obtenues par les modèles, sur un jeu de données indépendant
- **GO / NO-GO pour lancer une industrialisation**



La réalisation d'un projet I.A.

Industrialisation

C'est la phase de développement et d'intégration de la solution dans l'environnement informatique du client

--> c'est là où les profils Développeur sont particulièrement pertinents

- Réalisation des spécifications (environnement, interfaces, outils de visualisation...)
- Développement de la solution, tests unitaires, tests d'intégration...
- Recette par le client
- Mise en production

Outils employés :

- **Python, Scala, C++**
- **Gitlab-CI** : intégration continue
- **Docker**
- Bases de données SQL & NoSQL : PostGreSQL, MongoDB, CouchBase...
- IHM : Dash, Angular, React, Tableau, Grafana...



La réalisation d'un projet I.A.

Maintenance

Une fois la solution déployée, il faut **assurer sa maintenance** :

- Maintenance "technique" : corriger des anomalies techniques, faire des évolutions si nécessaires, gérer les montées de version des outils...
- Maintenance "I.A." : assurer le **maintien du niveau de performances des modèles algorithmiques** : **réentraînements réguliers**

La maintenance est parfois assurée par les équipes du client, une fois formées par les équipes Probayes



Offre d'emploi



Offre d'emploi

Data Scientist – Docteur en Sciences

Probayes a ouvert au recrutement une offre "Data Scientist – Docteur en Sciences", notamment dans le cadre de **projets subventionnés**, qui permettent de se former en Data Science sur plusieurs années.

Compétences requises :

- Analyse des données (statistiques, probabilités, visualisation de données)
- Programmation : Python, pandas, numpy
- Outils de travail en équipe : Git, Gitlab CI
- Français et anglais (oral et écrit)

Compétences appréciées :

- Première expérience en apprentissage automatique (sklearn, xgboost)
- Première expérience en Deep Learning (tensorflow, Pytorch)
- Connaissances en Big Data (Spark, Dask)
- Connaissances en bases de données (PostgreSQL, MySQL, MongoDB ...)
- Langages de programmation : C/C++, Java, Scala

Qualités requises :

- Vous êtes autonome, curieux(se) et rigoureux(se)
- Vous aimez travailler en équipe
- Vous avez une bonne communication orale et écrite

Offre d'emploi

Data Scientist – Docteur en Sciences

Informations complémentaires :

- Poste : CDI
- Lieu : Grenoble
- Date de début de contrat : dès que possible
- Candidature : Lettre de motivation + CV
- Candidature à adresser à : jobs@probayes.com
- Ce poste est ouvert, à compétences égales, aux personnes en situation de handicap

De jeunes docteurs en physique sont des candidats idéals !



Questions / Réponses



Ensemble, construisons vos solutions
d'Intelligence Artificielle

Merci pour votre attention!

Vincent Bonnivard

vincent.bonnivard@probayes.com

Alexandre Ghelfi

alexandre.ghelfi@probayes.com

www.probayes.com

 Probayes

 Probayes_IA

 Probayes