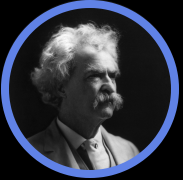


Cas d'étude: churn d'une clientèle de banque privée

Lombard Odier, Entretien Groupe Data Science

SAOUTER PIERRE, LE 20 JUIN 2025

Structure de la présentation



"The secret
to getting
ahead is
getting
started."

TWAIN

- ① **Churn: contexte et solution**
- ② **Conception d'un prototype**
- ③ **Résultats**
- ④ **Prochaines étapes**



Section 1

Churn: contexte et solution



Churn: contexte et solution

Stratégie de gestion de churn en banque privée

Pourquoi est-ce important?

Une pluralité de nouveaux acteurs

Les Intermédiaires Financiers Non Bancaires (IFM ou CGP en France), les Multi-Family Office (MFO), les Gestionnaires de Fortune Agréé, etc. Tous veulent une part du gâteau.

Une génération avec de nouvelles attentes

Rapport de Oliver Wyman et Morgan Stanley (2022): 66% des clients interagissent régulièrement avec des canaux digitaux pour la gestion de leur finance, y compris 40% des plus de 65 ans (x2 depuis 2019).

Un churn difficile à observer

Anticiper churn en banque privée souvent difficile: **Latent** (portefeuille progressivement déplacer sans fermer de compte) et **Silencieux** (clients restent polis, ne se plaint pas, mais est moins engager).

Stratégie de gestion de churn en banque privée

D'une posture réactive à pro-active

Une stratégie prédictive peut identifier les signes faibles

- baisse de fréquence de contact avec le banquier
 - réduction des actifs ou transferts sortants
 - désintérêt progressif pour les solutions d'investissement / utilisation moins importante des services
-

Personnalisation des actions préventives

- Mobilisation du banquier pour une relance personnalisée
- Proposition d'offre exclusive et de révision tarifaire
- Adapter le niveau de service / mettre en avant produit différentiants

Stratégie de gestion de churn en banque privée

D'une posture réactive à pro-active

Objectif

Le but n'est pas de proposer une automatisation à outrance mais d'anticiper les potentiels cas de churn et aider la banquier à réagir en lui proposant des outils adaptés.

L'humain reste la pierre angulaire de la relation client.

Stratégie de gestion de churn en banque privée

L'opportunité présentée par l'Intelligence Artificielle Générative

- **Les résultats des modèles prédictifs de Machine Learning ne sont pas faciles à communiquer**
 - Certaine réglementation demande de l'interprétabilité
 - L'incompréhension engendre la suspicion et une faible adoption
- **Les solutions d'IA générative offre de nouvelles opportunités pour présenter les résultats:**
 - Traductions des résultats de prédictions en "codes de langage"
 - Enrichissement des cas signalés avec informations internes et externes
 - Aide à la constitution de mesures réactives

Stratégie de gestion de churn en banque privée

Stratégie personnelle pour aborder ce cas d'étude

Fort de nos conversations préalables, l'objectif de ce travail est de montrer:

- Une approche orientée produit et métier
- La mise en place rapide d'un prototype (et avec peu de moyens)
- Comment un tel prototype pourrait être géré avec plus de moyens
- Les étapes vers une mise en production



Section 2

Conception d'un prototype

Une approche orientée produit

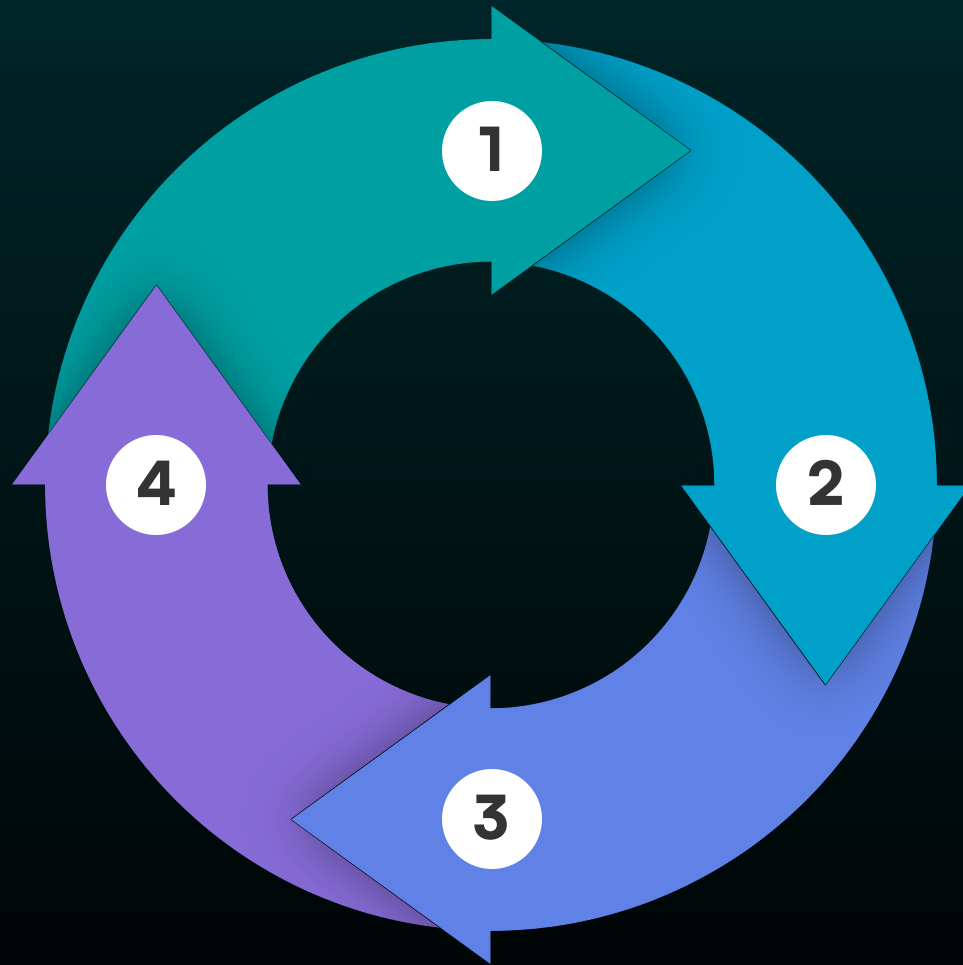
Nous concevons pour des utilisateurs métiers et non techniques

Nous répondons à un besoin humain et un objectif "business". En ce sens, notre solution doit se voir comme un produit qui:

- Répond aux besoins du métier
- Apporte un service à valeur ajoutée qui n'existe pas au sein de la banque
- Propose une expérience d'utilisation fluide et efficace (UX / UI)
- Est évolutif, répondant aux nouveaux besoins et réalités.

Prototype - Expérimentation itérative

En aucun cas synonyme d'un produit dysfonctionnel / inachevé



1

Produit Cibl 

D finition de la solution attendue par la m tier.
Co-cr ation avec le m tier.

2

Faisabilit 

Etude de faisabilit  (e.g. pr dictions du churn est possible avec la donn     disposition)

3

Prototype

Approche agile, peu de contraintes, limitations connues

4

Retour utilisateur

Le retour m tier permet d'it rer la cible et reprendre un cycle de d veloppement.

Quelques hypothèses pour le prototype

La phase de prototype requiert une certaine agilité de choix et d'exécution.

- **La phase de consultation avec le business a eu lieu: le problème est clair et nous avons une idée de la solution attendue.**
- **L'équipe data science a accès à la donnée nécessaire pour tester la performance d'un modèle prédictif de churn.**
- **Une étude préliminaire laisse penser que le churn peut être prédit avec une certaine performance (échantillon de labels, données historiques)**
- **L'équipe data science possède une certaine autonomie dans l'utilisation des solutions technologiques qui leur sont mises à disposition.**

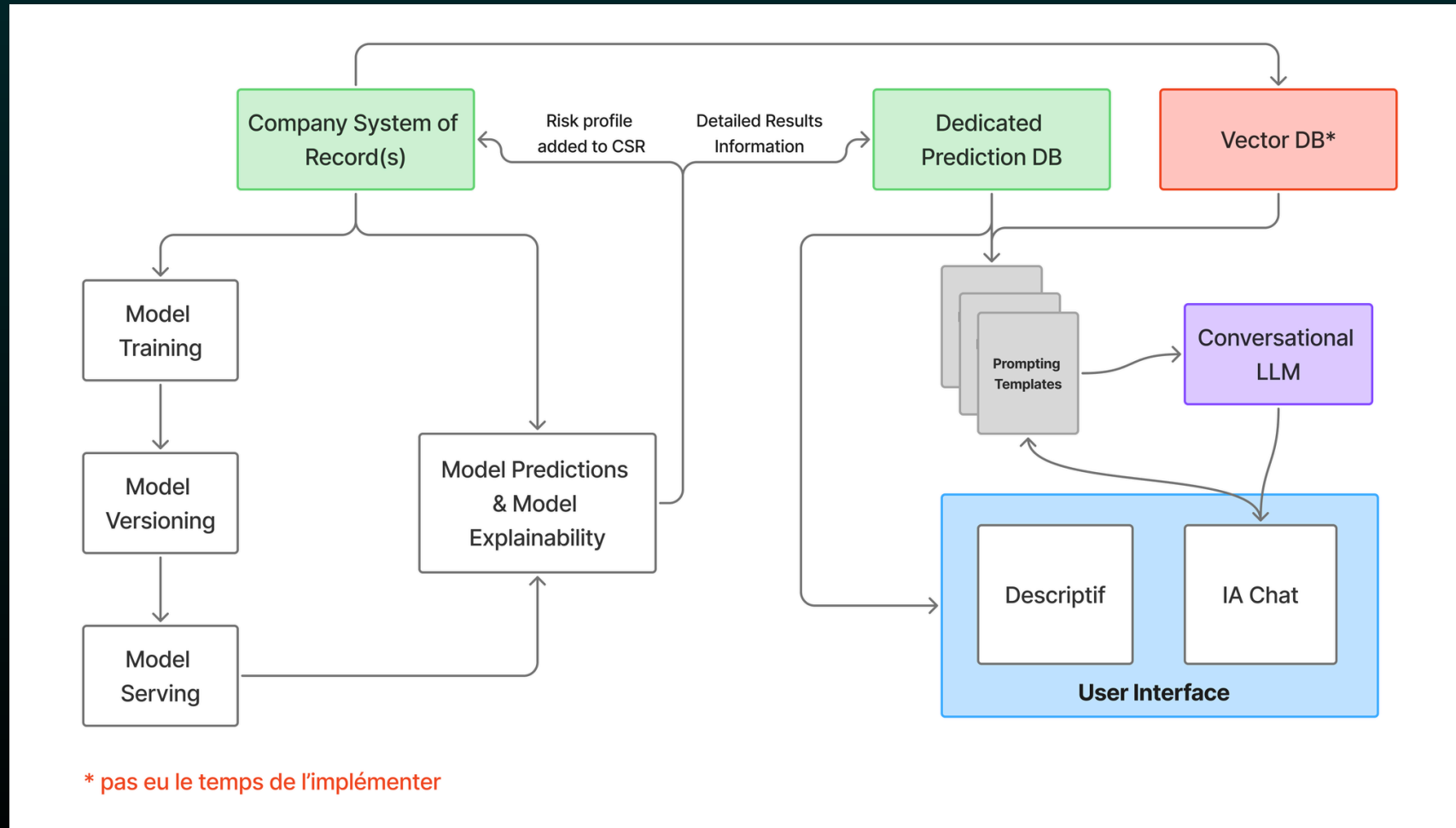
Quelques objectifs pour le prototype

Les attentes du métier - un assistant non-invasif pour comprendre et réagir

- Signaler les clients avec un avertissement simple : aucune probabilité affichée, niveau de risque de churn : faible, moyen, élevé.
- Fournir une explication des 3 principales variables pour justifier les raisons probables du churn du client.
- Traduire les variables en langage métier, avec une courte phrase ou sous forme de point clé.
- Fournir une suggestion d'action concrète basée sur les variables identifiées. (*Par exemple : "Client shows reduced interaction — consider scheduling a personal follow-up call."*)

Architecture de la solution proposée

Un prototype qui répond aux cas d'usage principaux



Outils utilisés pour le prototype

Des solutions Open Source (quasi que...) et un déploiement en local

- **Exploration de la donnée et ML**

- Ecosystème Tidyverse et TidyModels (parsnip, recipe, workflows, etc.)
- Archivage composants modèles dans pins versionnés dans répertoire local
- Modèle déployé sur API endpoint (plumber)

- **Sauvegarde de la donnée**

- Pins versionnés dans dossier local
- Base de donnée SQLite pour les résultats de prédictions

- **LLM**

- Librairie Posit Ellmer
- Ollama en local et Open AI API

- **Application Web**

- Framework R / Shiny
- Shinychat pour l'interface assistants
- CSS

- **Développement**

- Language R (librairies Posit compatibles R / Python et portable vers AWS, Azure)
- VSCode and Cursor
- Rapports générés avec Quarto
- GitHub Repository (GitHub Pages for static Quarto notebook analysis)
- MacBook Pro M1
- ChatGPT / Claude



Section 3

Résultats

Résultats

Résultats, codes et démonstration disponible en ligne

Répertoire GitHub avec l'ensemble des ressources

- https://github.com/principles-analytics/churn_prediction

Exploration des données et modélisation

- https://principles-analytics.github.io/churn_prediction/report_churn_prediction.html

Vidéo de démonstration

- <https://drive.proton.me/urls/MWBOBYHKJC#RTU3BRLEZFP8>

Prédiction de churn

Analyse exploratoire de l'échantillon partagé (voir le [rapport](#))

- **Echantillon propre sans valeur manquantes**
- **Echantillon présente 1000 clients, dont 190 ont churné (~19%)**
- **Les caractéristiques de l'échantillon semble limiter le potentiel de l'analyse**
 - Matrice de corrélation ne montre aucune co-linéarité
 - Distribution de chaque variable quasi uniforme
 - Pouvoir prédictif inexistant
 - Variables clé d'une analyse de churn est manquante: notion de temporalité
- **Explications possible**
 - Données probablement générées artificiellement / sans lien avec la réalité
 - Les variables avec pouvoir prédictif sont manquantes

Prédiction de churn

Jouons le jeu de l'échantillon de données partagé

Modèles testés pour la prédiction

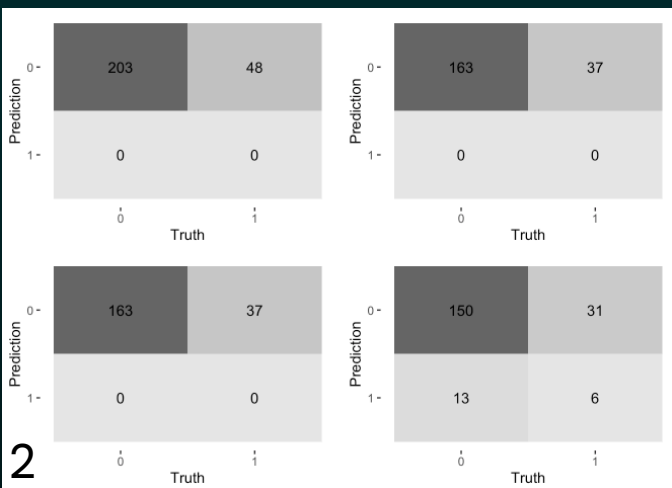
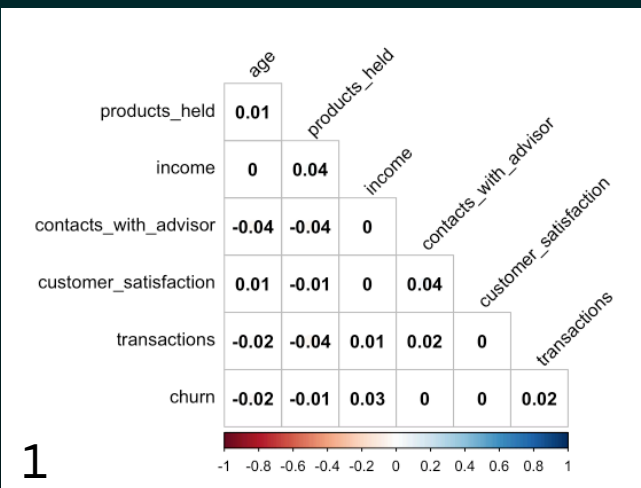
- **Régression logistique** : première étape simple, facile à interpréter, à expliquer et à « vendre ». Moins performante avec de petits échantillons ou des effets non-linéaires importants.
- **Régression logistique pénalisée**: sélection des prédicteurs (features) pendant l'entraînement au travers d'une méthode pénalité (méthode de Lasso).
- **Random Forests**: entraînement en parallèle de plusieurs arbres de décision sur des variantes échantillonnées de la donnée d'entraînement.
- **Boosted trees (XGboost)** entraînement séquentiel d'arbres de décision venant corriger / améliorer la performance du précédent cycle d'entraînement.

Prédiction de churn

Concepts généraux de l'approche

- **Sur la base de l'échantillon test:**
 - training set (60%-89%)
 - validation set (20%) utilisé pour les modèles avec tuning de hyperparameter
 - test set (20%) utilisé pour qualifier la performance finale du modèle
- **Performance des modèles**
 - Matrice de confusion
 - Accuracy, Sensitivity et Specificity
 - Courbes de RoC
 - Courbes de Lift
- **Interprétation / "Explainability"**
 - Analyse de coefficients (pour les modèles linéaires)
 - Analyse de SHAP (effet global et local, valeurs SHAP)

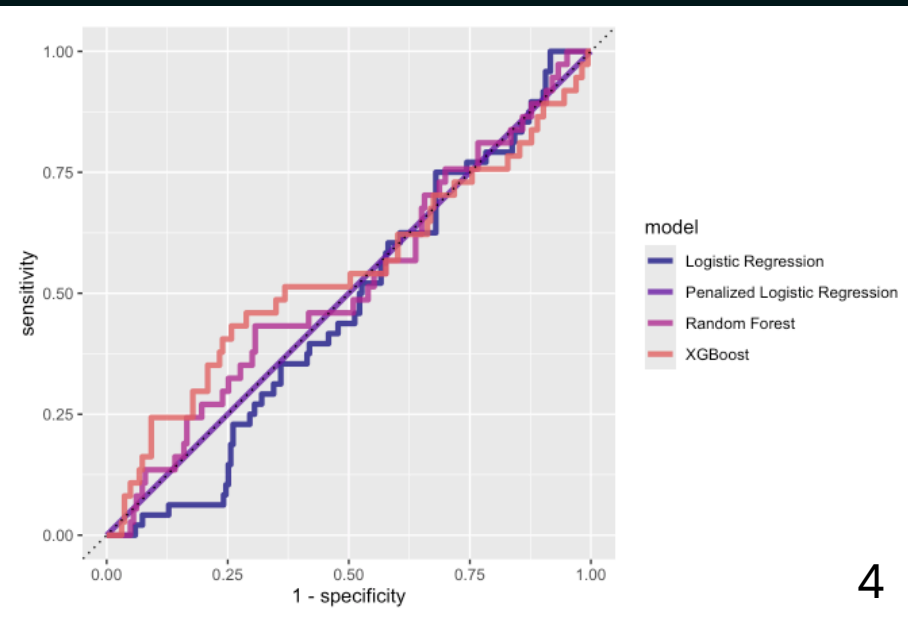
Quelques résultats



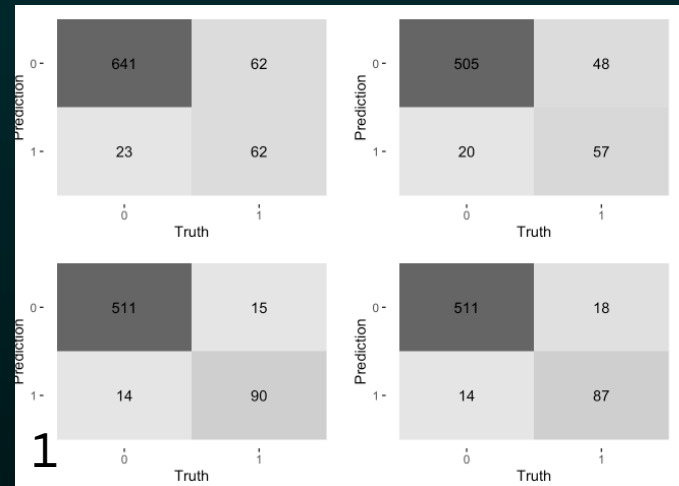
- 1. matrice de correlation
- 2. matrices de confusion
- 3. accuracy, sensitivity, etc.
- 4. Courbes de RoC

| Model | Performance Metric | Estimated Performance |
|-------------------------------|--------------------|-----------------------|
| Logistic Regression | accuracy | 80.88 |
| Logistic Regression | kap | 0.00 |
| Logistic Regression | sens | 0.00 |
| Logistic Regression | spec | 100.00 |
| Penalized Logistic Regression | accuracy | 81.50 |
| Penalized Logistic Regression | kap | 0.00 |
| Penalized Logistic Regression | sens | 0.00 |
| Penalized Logistic Regression | spec | 100.00 |
| Random Forest | accuracy | 81.50 |
| Random Forest | kap | 0.00 |
| Random Forest | sens | 0.00 |
| Random Forest | spec | 100.00 |
| XGBoost | accuracy | 78.00 |
| XGBoost | kap | 10.15 |
| XGBoost | sens | 16.22 |
| XGBoost | spec | 92.02 |

3



Un exemple avec un échantillon Kaggle

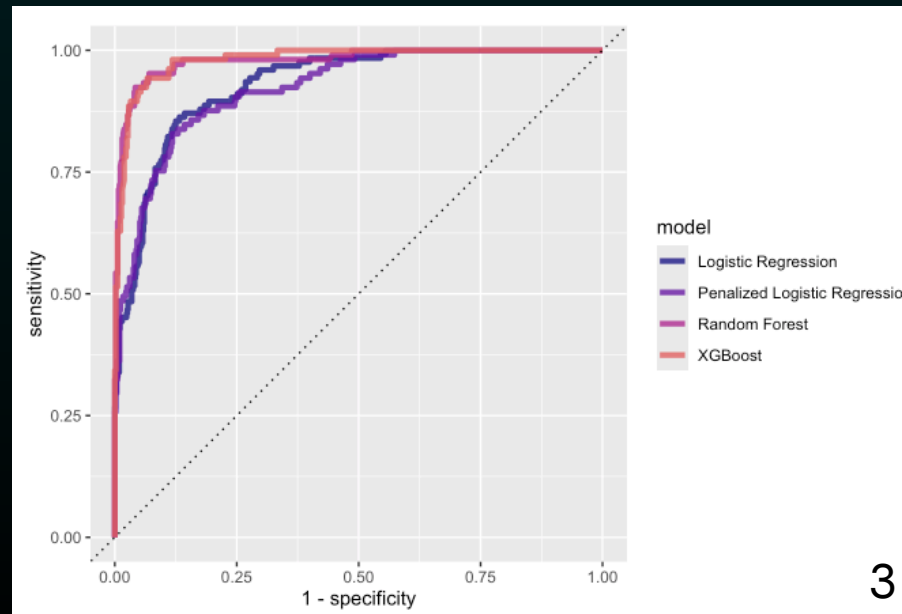


- 1. matrices de confusion
- 2. accuracy, sensitivity, etc.
- 3. Courbes de RoC

Echantillon churn d'une société de télécommunication iranienne.

| Model | Performance Metric | Estimated Performance |
|-------------------------------|--------------------|-----------------------|
| Logistic Regression | accuracy | 89.21 |
| Logistic Regression | kap | 53.36 |
| Logistic Regression | sens | 50.00 |
| Logistic Regression | spec | 96.54 |
| Penalized Logistic Regression | accuracy | 89.21 |
| Penalized Logistic Regression | kap | 56.50 |
| Penalized Logistic Regression | sens | 54.29 |
| Penalized Logistic Regression | spec | 96.19 |
| Random Forest | accuracy | 95.40 |
| Random Forest | kap | 83.37 |
| Random Forest | sens | 85.71 |
| Random Forest | spec | 97.33 |
| XGBoost | accuracy | 94.92 |
| XGBoost | kap | 81.43 |
| XGBoost | sens | 82.86 |
| XGBoost | spec | 97.33 |

2



Prédiction de churn

Sélection d'un modèle pour la mise en production

Etant donné les résultats, nous avons arbitrairement sélectionné le modèle XGBoost pour la suite (challenge de l'interprétabilité)

- **Versionnage du un modèle**
 - pins en local (substitution vers Azure Blob ou AWS S3 en une ligne de code)
 - nous sauvons: ingrédients du fit, liste des features, testing data set
- **Déploiement du modèle**
 - déploiement sur un API endpoint (en local) en utilisant plumber
 - au final, le cas d'usage ne nécessite pas des prédiction en temps réels. Nous réalisons des prédictions à intervalle régulier que nous sauvons en base.

Prédiction de churn

Modèle accessible depuis un endpoint

The screenshot displays the Vetiver 0.2.5 web interface in a browser. The address bar shows the URL `127.0.0.1:8080/_docs_/#post-/predict`. The interface includes a sidebar with a search bar and a list of API endpoints under the heading "churn_model_xgboost model API". The main panel is titled "REQUEST" and shows the "REQUEST BODY" as `application/json`. An "EXAMPLE" tab is selected, displaying a JSON object with fields: `"client_id": 0`, `"transactions": 0`, `"contacts_with_advisor": 0`, `"customer_satisfaction": 0`, `"products_held": 0`, `"age": 0`, and `"income": 0`. Below the request section, the "API Server" is listed as `http://127.0.0.1:8080/` with "Authentication Not Required". The "Response Status" is `OK:200` and it took `110 milliseconds`. A "RESPONSE" tab is selected, showing a JSON object with `".pred_class": "0"`. Buttons for "FILL EXAMPLE", "CLEAR", "TRY", "CLEAR RESPONSE", and "Copy" are visible.

vetiver 0.2.5 `http://127.0.0.1:8080/openapi.json` LOCAL JSON FILE

Filter SEARCH

churn_model_xgboost model API

OPERATIONS

- Health check
- Get all metadata of pinned vetiver model
- Get input data prototype for vetiver model
- Return predictions from model using 7 features

REQUEST

REQUEST BODY `application/json`

EXAMPLE SCHEMA

```
[
  {
    "client_id": 0,
    "transactions": 0,
    "contacts_with_advisor": 0,
    "customer_satisfaction": 0,
    "products_held": 0,
    "age": 0,
    "income": 0
  }
]
```

API Server `http://127.0.0.1:8080/`
Authentication Not Required

Response Status: `OK:200`
Took 110 milliseconds

RESPONSE RESPONSE HEADERS CURL

```
[
  {
    ".pred_class": "0"
  }
]
```

Copy

FILL EXAMPLE CLEAR TRY

CLEAR RESPONSE

Assistants IA

Un assistant pour aider à comprendre le churn (proof-of-concept)

- **Création de template de prompt**

- Template "`interprétation du churn`": l'assistant est nourri des informations liés au client et à la prédiction et utilise le template et les exemples pour aider le banquier à comprendre les raisons du risque en langage simple.
- Template "`analyse du client`": l'assistant aide le banquier à sonder l'historique des relations du client avec la banque pour mieux comprendre.
- Template "`recommandations`": élaboration de recommandations pour prendre des mesures à l'égard des cas de churn suspectés (actions, draft the mail).

- **Dans le futur**

- Approche RAG (une partie du contexte est dynamiquement recherchée dans db vectorielle)
- Assistant à Agent (MCP pour rendre l'agent autonome dans le recherche d'information)

Données synthétiques

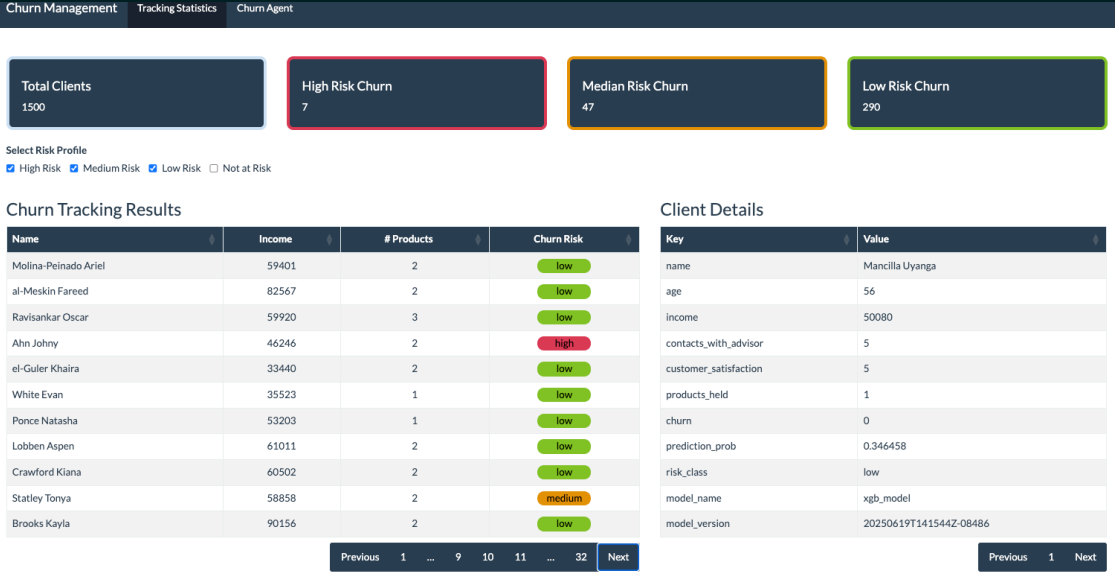
Pour permettre une expérience plus réaliste du produit (voir [ici](#))

- **Création d'un échantillon de données synthétiques**
 - 5000 clients avec prénom, nom, temps passé en tant que client
 - Prédiction de churn basé sur le modèle en production
 - Profils de risque: faible, moyen, fort, sans-risque
 - Explication des prédictions avec valeur de SHAP
 - Création de code de langage pour chaque variables
 - ex.: transactions = "Significant change in transaction frequency"
- **Sauvegarde dans SQLite locale accessible par l'app**

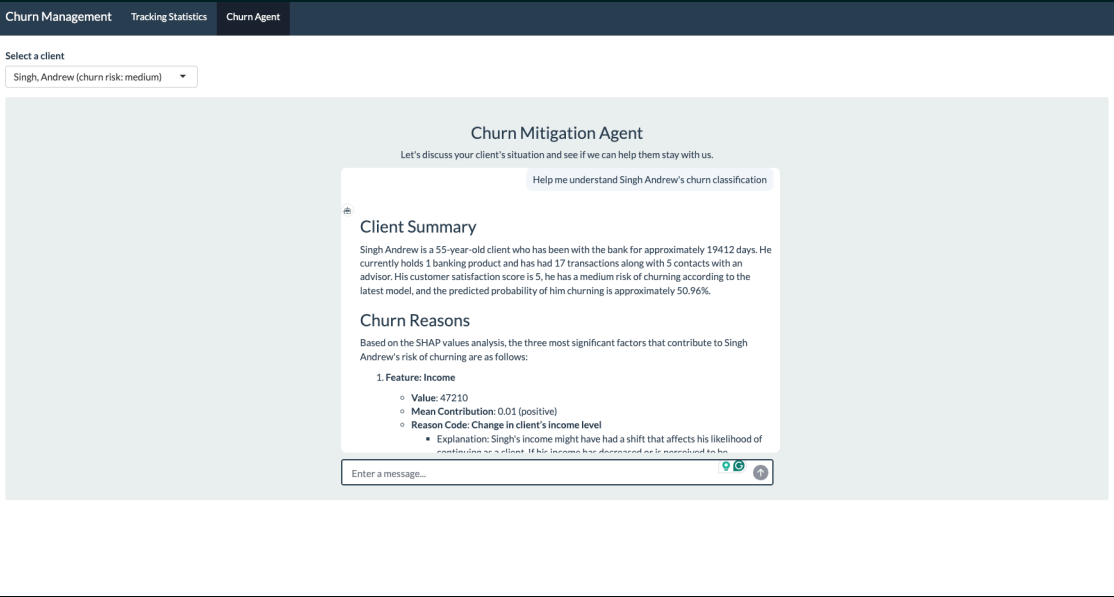
Le produit final

La UI / UX s'est pas mon fort mais le produit fonctionne! (voir vidéo)

Dashboard de monitoring



Assistant IA





Section 5

Prochaines étapes

Ce qu'il manque au prototype

Avec un peu plus de temps et un plus large choix d'outils

- **Question de temps**

- Nettoyage du code
- Gestion de l'environnement et conteneurisation (docker)
- Approche RAG
- Observabilité de la performance du modèle et ré-entraînement
- Template de prompt améliorés (base d'exemples solides)
- Stratégie d'évaluation des assistants LLM

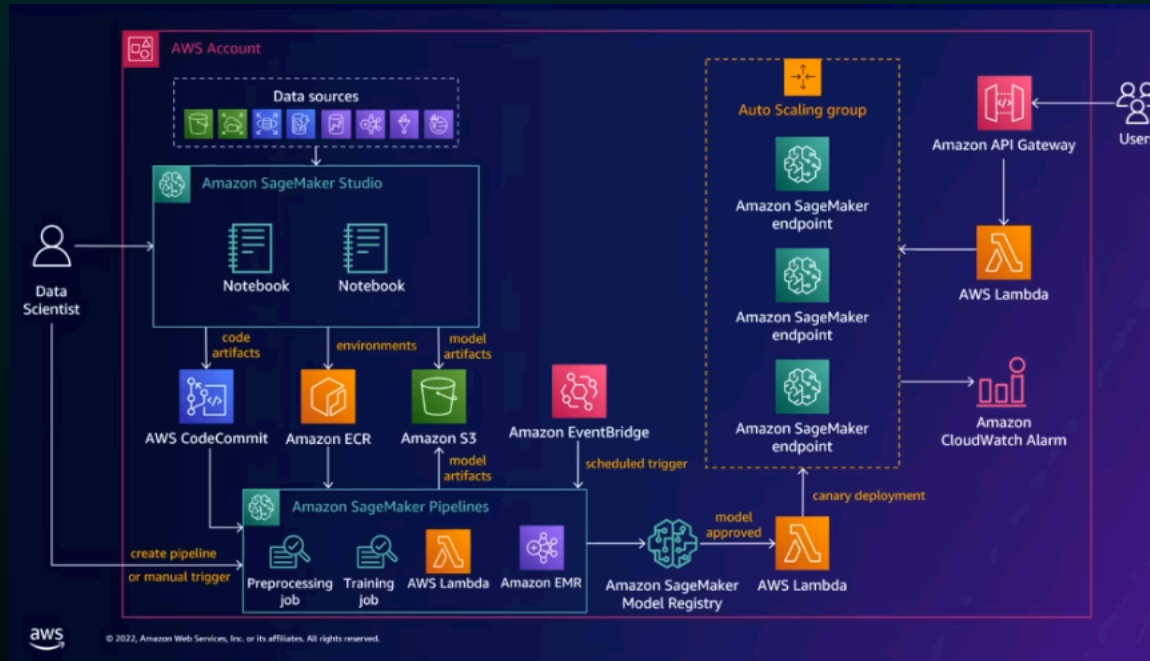
- **Question d'outils / de budgets**

- Développement sur un environnement type AWS
- Orchestration des systèmes
- Architecture MLOps
- Permet de mieux saisir les conditions d'une mise en production

Exemple d'un prototype construit dans le cloud

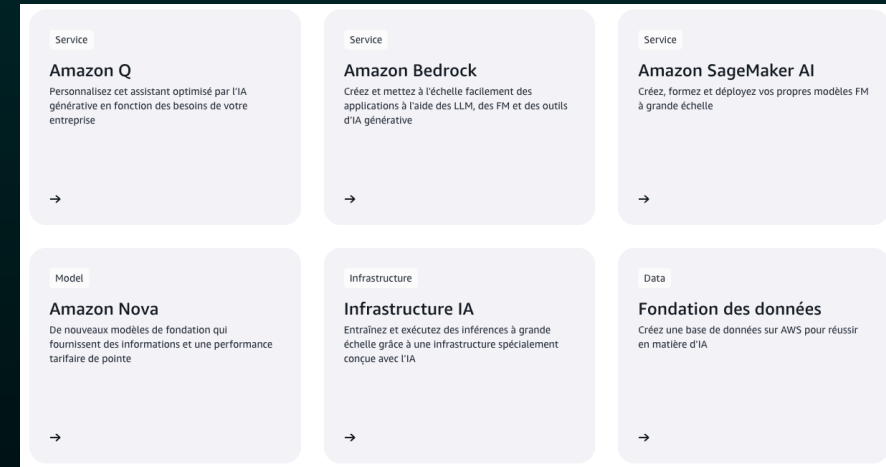
Small MLOps / AIOps Architecture on AWS (can be on Azure, DataBricks, ...)

Architecture MLOps



<https://aws.amazon.com/fr/awstv/watch/Oed9d2d9b05/>

Out-of-the box AI services



Product

Le prototype dans la réalité

Conditions de ce projet mené en entreprise

Approche Projet Agile

1. Un chef de projet dédié supervise / participe aux développements et maintient le lien avec le business
2. Petite équipe avec des compétences polyvalentes travaillent sur le projet en mode agile.

Dimension IA Générative

1. Contraintes d'entreprise: modèles disponibles, réglementation, Eval
2. Intégration avec les autres systèmes (Agents)
3. Données et bases de données accessibles

Machine Learning

1. Données accessible pour entraîner le modèle.
2. Evaluation de la performance du modèle
3. Monitoring de performance en production, etc^.

Dimension Produit et Gouvernance

1. Compétence spécifique autour de la question de la UX / UI de la solution.
2. Respect de l'identité de l'entreprise et l'alignement avec les autres solutions
3. Respect des règles de gouvernance interne

Dimension Architecture et Systèmes

1. Importance d'une compétence DevOps pour l'aspect déploiement
2. Question de la communication avec les systèmes tierces tels que les clients Mail, les bases de données, des sources externes.

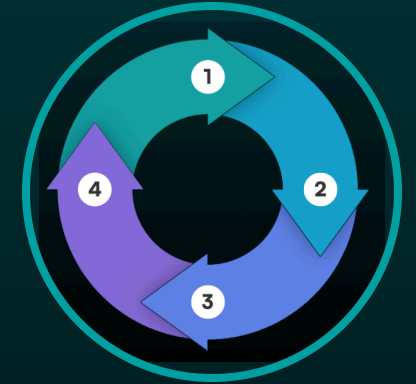
Questions autour de la validation du prototype

Plusieurs questions essentielles se posent avant d'aller vers la production

- Est-il possible de prédire le churn avec une précision suffisante
- La qualité des agents IA est-elle suffisante et sans risque
- Quel est le coût et l'effort pour atteindre une performance et une qualité suffisante?
- Le produit est-il utilisable et utilisé par le métier
- Ce produit est-il durable et cost-effective
- Ce produit répond-t-il aux exigences de gouvernance interne

De l'expérimentation à la production

Un cycle de validation et d'itérations vers le produit final.



Produit Cibl 

D finition de la solution attendue par la m tier.
Objectifs du prototype.

Agile, contraintes limit es.
Approche dev. produit.

Validation

Retour d'exp rience par la m tier.
Validation par un comit  projet.

Produit cible est pr cis , scope V1 d fini. Dev produit V1

Production

Approche projet IT: DEV(OPS), ML Engineers, MLOps
Cycle produit

Prototype

V1 Product

Merci pour votre attention

Pierre Saouter