

# RAPPORT DE CONCEPTION ET D'ANALYSE DE L'HOSPITALIÈRE PITIÉ SALPETRIÈRE

Lounes TAIBI - Kenza Mona EL HITARY - Hope BOME - Aissatou BLONDIN - Khady BA

## *Optimisation des Flux par l'IA*

### 1. Présentation du Prototype et Méthodologie

#### 1.1. Études Préalables et Sources de Données Publiques

Le développement du prototype Salpêtrière-Predict ne repose pas uniquement sur des données internes, mais sur une intégration des standards de santé publique français. Pour calibrer notre modèle, nous avons exploité les ressources de la DREES (Direction de la recherche, des études, de l'évaluation et des statistiques) :

- Capacité des lits et typologie de soins : L'analyse du dataset Lits de réanimation, SI et SC a permis de définir les seuils de saturation par service et de comprendre la distribution de l'offre de soins critiques. [lien](#)
- Modélisation des parcours complexes : L'utilisation des données sur les patients hospitalisés pour gestes auto-infligés a servi à simuler des flux spécifiques nécessitant une surveillance accrue et des ressources humaines spécialisées. [lien](#)

#### 1.2. Méthodologie et Management de Projet

Le projet a été mené selon une approche Agile hybride, garantissant une livraison itérative tout en respectant les contraintes hospitalières :

1. **Phase de Cadrage** : Identifier les points de douleur (saturation le lundi matin, manque de visibilité sur les sorties).
2. **Fonctionnalités du Prototype (Dashboard)** : Le prototype développé ("Hospital Command Center") est une application web interactive destinée aux directions hospitalières et aux gestionnaires de lits (Bed Managers). Il s'articule autour de quatre fonctionnalités majeures :
  - Météo Hospitalière (Prévision J+1) : Un système d'alerte immédiat (Vert/Jaune/Rouge) basé sur la prédiction de flux à 24h, couplé à des consignes opérationnelles automatiques (ex : "Activation Plan Blanc").
  - Exploration Temporelle (2018-2026) : Une interface de navigation permettant d'analyser les tendances historiques et de se projeter dans le futur via des données générées simulant des tendances long terme.
  - Simulateur de Scénarios (Stress Test) : Un moteur "What-If" permettant de modifier manuellement les paramètres critiques (Choc de flux épidémique ou Choc RH/Grève) pour observer la résilience de l'hôpital en temps réel.
  - Gestion Opérationnelle (Vue Bed Manager) : Un tableau de bord détaillé par service (Urgences, Chirurgie, etc.) calculant la capacité nette en temps réel (Lits Totaux - Lits Fermés pour cause RH).
3. **Management des Risques & Acceptabilité** : Pour favoriser l'adoption par les soignants, nous avons opté pour une approche "White Box" (modèles explicables)

plutôt qu'une boîte noire algorithmique. Chaque prédiction est un outil d'aide, laissant le dernier mot à l'expertise humaine.

## 2. Analyse des Tendances et Diagnostic Stratégique

### 2.1 Identification des Périodes Critiques

L'analyse des séries temporelles révèle trois cycles de tension superposés :

- Le Cycle Saisonnier (L'Hiver) : Une augmentation structurelle du flux de +15% à +20% durant les mois de décembre, Janvier et février, corrélée aux épidémies hivernales (Grippe, Bronchiolite).
- Le Cycle Hebdomadaire (L'effet lundi) : Une asymétrie marquée de la semaine. Le Lundi est systématiquement le jour le plus chargé (rattrapage du week-end, ouverture de la médecine de ville), tandis que le dimanche présente le flux le plus bas mais souvent l'effectif le plus réduit.
- Les Chocs Exogènes (Type Covid) : Des ruptures de tendance brutales (pics ou creux massifs) qui sortent des distributions normales et nécessitent des modèles robustes aux "outliers".

### 2.2 Évaluation des Stratégies Actuelles

La gestion actuelle, souvent basée sur la moyenne historique ou l'intuition, montre ses limites :

- Réactivité vs Proactivité : Les renforts sont souvent appelés *après* que la saturation soit constatée (mode pompier), ce qui coûte plus cher et épuise les équipes.
- L'Angle Mort RH : Les stratégies actuelles se focalisent sur les lits physiques. Or, notre analyse montre que la saturation est souvent causée par la fermeture de lits due à l'absentéisme, et non par un manque de murs.

## 3. Analyse Statistique, Justification de la Modélisation

Note Théorique : Un modèle statistique est une représentation mathématique décrivant la génération des données observées. Il formule des hypothèses sur la distribution des variables et leurs relations. L'objectif est d'expliquer le passé pour projeter le futur, tout en quantifiant l'incertitude associée à chaque prédiction.

### 3.1 Justification des Choix de Visualisation

- Histogrammes & Distributions (KDE) : Utilisés pour visualiser la distribution des Âges et des Durées de Séjour (LOS).
- *Justification Statistique* : Vérifier la normalité des distributions. La découverte d'une distribution "Log-Normale" sur la durée de séjour (longue traîne à droite) justifie l'utilisation de la médiane plutôt que la moyenne pour éviter les biais dus aux valeurs extrêmes.
- Boxplots (Boîtes à moustaches) : Utilisés pour croiser Gravité (CCMU) et Durée de Séjour.

- *Justification Statistique* : Détecter les outliers (valeurs aberrantes). Cela a permis de prouver que la gravité médicale n'est pas toujours corrélée à la durée de séjour (des patients peu graves peuvent rester longtemps faute de solution sociale).
- Matrices de Corrélation (Heatmaps) : Utilisées pour analyser les liens entre Variables (ex : Absentéisme et Saturation).
- *Justification Statistique* : Mesurer la force des dépendances linéaires (Coefficient de Pearson) pour sélectionner les variables pertinentes (Feature Selection) pour le modèle prédictif.

### 3.2 Modèles Statistiques Sous-jacents

Nous avons formulé les hypothèses suivantes :

1. Distribution : Le flux journalier de patients suit approximativement une loi de Poisson (événements rares et indépendants) pour les arrivées, mais peut être approximé par une loi Normale pour les grands nombres (Théorème Central Limite).
2. Stationnarité : Les séries temporelles ne sont pas stationnaires (elles ont une tendance et une saisonnalité). Il a fallu les décomposer pour les modéliser.

### 4. Les Modèles de Prédiction

#### 1. Modèle 1 : Régression Linéaire (Baseline)

Classique via l'équation linéaire  $\$Y = aX + b\$$  (où  $\$Y\$$  est le flux prédit,  $\$X\$$  le jour de la semaine,  $\$a\$$  le coefficient d'impact et  $\$b\$$  la constante). Malgré une erreur moyenne (**MAE**) de **~17 patients**, ce modèle s'avère insuffisant pour le pilotage de crise car il échoue à capturer les **non-linéarités**, "**Lundi + Hiver + Grève**" génère une saturation bien plus brutale que la simple somme de ces trois éléments. Ce constat justifie l'abandon de la régression simple au profit du **Machine Learning (XGBoost)**, capable de modéliser ces interactions complexes et d'anticiper les ruptures de charge.

#### 2. Modèle 2 : XGBoost Regressor (Retenu)

Algorithme de "Gradient Boosting" basée sur des arbres de décision construits séquentiellement pour corriger les erreurs des modèles précédents. Contrairement à la régression classique, XGBoost capture parfaitement les **effets de seuil**, comme l'explosion de la saturation dès que l'absentéisme franchit la barre des 15 %. Cette **approche hybride** permet de croiser efficacement des variables catégorielles (jours de la semaine) et continues (taux d'absentéisme), tout en étant moins sensible aux valeurs aberrantes. Grâce à ce modèle, nous avons réduit l'erreur moyenne (**MAE**) à **~13,5 patients**, soit un gain de précision de **23 %** par rapport à la baseline. Enfin, l'intégration de l'analyse **SHAP** a permis de hiérarchiser les facteurs d'influence, révélant que la variable "Hiver" est le prédicteur dominant, suivie de près par la saisonnalité hebdomadaire et le niveau des effectifs présents.

### Impact et Perspectives du MVP

L'utilisation de Pitié-Salpêtrière transforme la gestion de flux :

1. Impact Humain : Réduction du stress des équipes par une meilleure anticipation des renforts.

2. Impact Opérationnel : Fluidification du parcours patient et réduction des temps d'attente aux urgences de l'ordre de 15%.
3. Évolutivité : Intégration prochaine de données en temps réel via le Dossier Patient Informatisé (DPI).