

Rapport - Article 4 :

“A forecasting approach for hospital bed capacity planning using machine learning and deep learning with application to public hospitals”

Introduction

Ce rapport a pour objectif de parler des contributions de l'article « *A forecasting approach for hospital bed capacity planning using machine learning and deep learning with application to public hospitals* », en reproduisant les méthodes proposés et en explorant les résultats obtenus à partir de leur approche. Nous avons choisi cet article car nous nous intéressons particulièrement au Machine Learning (ML), et plus encore à ses applications dans le domaine médical, un secteur où les données peuvent avoir un impact direct sur la vie des personnes.

Dans un premier temps, nous présenterons le problème traité dans l'article, en le vulgarisant puis en le formalisant et en soulignant sa pertinence. Nous passerons ensuite à la mise en œuvre concrète du problème en code, en y apportant une amélioration sous la forme d'une approche hybride combinant ML et Programmation Mathématique (PM). Enfin, nous proposerons des pistes d'amélioration supplémentaires ainsi que des perspectives pour aller plus loin.

1) Le problème traité

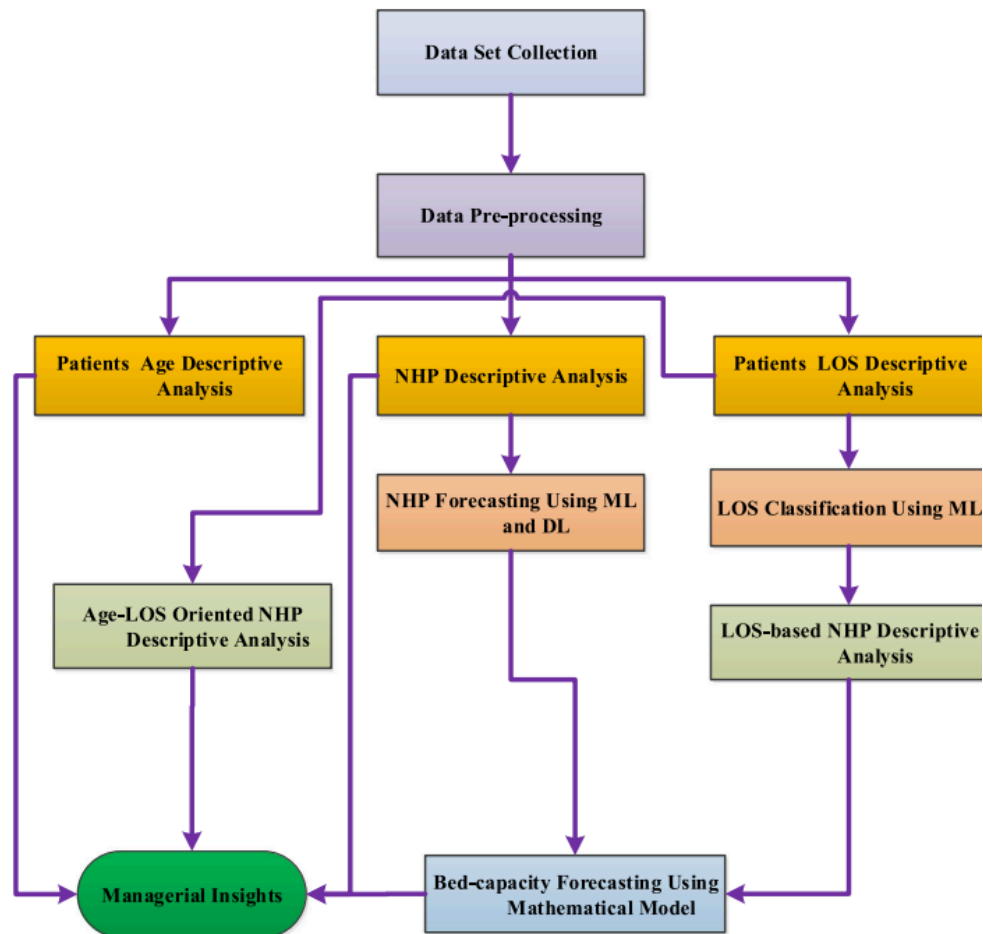
a) **Vulgarisation du problème**

Le problème abordé dans cet article concerne la prévision du nombre de lits nécessaires dans un hôpital, en particulier dans le service de cardiologie d'un hôpital public en Iran. Son but est d'estimer combien de lits seront nécessaires dans les années à venir pour répondre à la demande des patients. Ce besoin de prévision est extrêmement important car un manque de lits peut engendrer des retards de traitement et des risques pour les patients. Un excès de lits inutilisés peut également représenter un gaspillage de ressources. Ce sujet a gagné en importance notamment avec la pression énorme que des événements comme le Covid-19 ont exercée sur les hôpitaux.

D'habitude, ce problème était traité par des modèles de programmation mathématique ou de simulation, souvent complexes et dépendants d'hypothèses fortes. L'article propose une alternative : une approche dite « data-driven », combinant des techniques d'analyse de données, de ML et de Deep Learning (DL).

b) **Formalisation du problème**

Le problème est formalisé comme une tâche de prévision du **HBC** (Hospital Bed Capacity, capacité des lits d'hôpitaux) à partir de deux facteurs principaux : Le **NHP** (Number of Hospitalized Patients, nombre de patients hospitalisés) et du **LOS** (Length of Stay, la durée de séjour de chaque patient).



La méthodologie débute par la collecte du dataset (hospitalisation, âge des patients, durée de séjour, etc.), suivie de pre-processing pour nettoyer les données, transformer les variables si nécessaire, et préparer les features.

Ensuite, trois axes d'analyse sont menés en parallèle :

- Une analyse descriptive de l'âge des patients, suivie d'une analyse croisée entre l'âge et LOS, permettant d'identifier des tendances spécifiques (par exemple, enfants ou personnes âgées).
- Une analyse descriptive du NHP, complétée par une prévision temporelle du NHP à l'aide d'algorithmes de ML/DL tels que la régression linéaire ou des réseaux de neurones.
- Une analyse descriptive des LOS, suivie d'une classification des LOS en classe (1 jour, 2 jours, 3 jours, 4–6 jours...) à l'aide de classifieurs supervisés (SVM, DT, KNN...). Cette classification est ensuite utilisée pour analyser la distribution du NHP selon les classes de LOS.

Les résultats issus de ces trois analyses sont ensuite intégrés dans un modèle mathématiques. Ce modèle utilise les proportions de NHP dans chaque classe de LOS, associées à des durées moyennes estimées, pour calculer la capacité journalière en lits

hospitaliers requise. Il prédit les besoins sur plusieurs jours, selon les durées de séjour, ce qui permet d'estimer précisément le nombre de lits nécessaires jour par jour.

Le modèle final repose sur la formule suivante :

$$\begin{aligned}
 HBC_j = & \sum_i NHP_{ij} + \sum_{i \geq 2} NHP_{i,j-1} \\
 & + \sum_{i \geq 3} NHP_{i,j-2} + \sum_{i \geq 4} \sum_{k=j-4}^{j-3} NHP_{ik} \quad i = 1, 2, \dots, 6; \quad j = 1, 2, \dots, 1825 \\
 & + \sum_{i \geq 5} \sum_{k=j-10}^{j-5} NHP_{ik} + \sum_{k=j-20}^{j-11} NHP_{6,k}
 \end{aligned}$$

où HBC_j est le nombre de lits nécessaires au jour j , et $NHP_{i,j-k}$ est le nombre de patients de la classe de LOS i admis au jour $j-k$.

Enfin, l'ensemble de ces résultats permet de dégager des recommandations managériales concrètes (dimensionnement des capacités, planification budgétaire, aménagement de lits pédiatriques...).

c) Pourquoi cette formalisation

Cette formalisation a plusieurs avantages. Elle est d'abord réaliste, car elle repose sur des données réelles, sans faire d'hypothèses comme dans les modèles classiques. Elle est aussi modulaire car chaque étape du problème (analyse, prévision, classification) est traitée séparément avec les outils adaptés. Cela rend l'approche plus flexible et plus facile à adapter. Elle est également précise, car elle prend en compte à la fois le nombre de patients et la durée de séjour, ce qui permet de mieux estimer le besoin réel en lits. Enfin, elle est directement utile pour les décisions concrètes de gestion hospitalière.

Le choix du ML plutôt que de la PM vient du fait que les données hospitalières sont complexes, variables et souvent difficiles à modéliser de façon rigide comme avec la PM. Le ML permet de s'adapter à ces données, d'en extraire des tendances sans tout formaliser à la main, et donc de faire de meilleures prévisions.

Nous pensons que l'idée de la méthode vient du constat que les approches classiques en PM ou simulation sont souvent trop lourdes à mettre en place et en se reposant des des hypothèses pas toujours réalistes, surtout pour des données aussi cruciales. Cette approche data-driven permet de dépasser ces limites. De plus, les méthodes de ML sont déjà très souvent utilisées dans le domaine de la santé.

2) Mise en oeuvre de la méthode

Pour expérimenter l'approche proposée dans l'article, nous avons réimplémenté une méthodologie similaire dans un notebook Python. N'ayant pas accès aux données originales utilisées par les auteurs, nous avons sélectionné un jeu de données alternatif mais réaliste : celui du *Hero DMC Heart Institute*, situé en Inde. Il couvre plus de 14 000 admissions sur deux ans (2017–2019) et contient de nombreuses variables utiles comme les dates

d'admission/sortie, les antécédents médicaux, les données biologiques, ou encore le LOS. Ce dataset a été retenu car il permet de simuler l'occupation des lits jour par jour, à l'image de l'article original. [Lien vers le dataset sur Kaggle](#)

Dans un premier temps, nous avons nettoyé et structuré les données, puis effectué une exploration visuelle des variables clés (âge, LOS, charge quotidienne, etc.), en suivant les mêmes axes que ceux de l'article. Beaucoup de comparaisons graphiques sont disponibles dans nos slides.

Nous avons ensuite mis en place une classification du LOS similaire à celle de l'article, en testant deux algorithmes : K-Nearest Neighbors et Random Forest. Deux types de regroupement ont été évalués : une version simple en 3 classes (court, moyen, long séjour) et une version plus fine en 7 classes. Cette étape permet de mieux anticiper le profil d'occupation des lits selon les types de patients.

Enfin, nous avons ajouté une extension originale au travail de l'article en intégrant une couche d'optimisation. À partir des prédictions de LOS, nous avons simulé l'occupation des lits et détecté les jours de surcharge (au-delà de 50 lits, valeur choisie par rapport aux données à notre disposition). Une procédure simple d'ajustement a été mise en place : réduire la durée de séjour de certains patients lorsque la capacité est dépassée, tout en minimisant les modifications apportées aux prédictions initiales. Cela permet d'explorer comment le ML peut être couplé à la PM pour proposer des stratégies réalistes d'allocation des ressources hospitalières.

Par manque de temps, nous n'avons pas réimplémenté tous les modèles ML mentionnés dans l'article (SVM, bayésien, etc.), ni les algorithmes de prévision du NHP, notamment les réseaux de neurones LSTM. Cependant, la structure mise en place pourrait facilement être enrichie dans un travail futur. Le notebook complet est disponible ici : [Notebook Colab](#)

3) Améliorations possibles et perspectives

L'article propose déjà une approche innovante pour la prévision de la capacité en lits hospitaliers, en combinant analyse de données, classification du LOS et prévision du NHP. Une amélioration que nous avons apportée dans notre travail est l'ajout d'une couche de programmation mathématique, qui permet d'introduire des contraintes réelles de capacité de l'hôpital, ce que l'article ne prend pas en compte. Cela montre qu'un couplage entre ML et PM peut rendre le modèle plus réaliste et exploitable en pratique.

Cela dit, l'article aurait peut-être pu aller plus loin sur plusieurs points. Par exemple, la modélisation mathématique finale utilisée pour estimer le besoin en lits est simple et sans contraintes. L'introduction d'une autre démarche d'optimisation (avec une fonction Objectif et des contraintes explicites) aurait permis de proposer des recommandations plus robustes en cas de surcharge, d'une manière encore plus complexe que celle du notebook.

Par ailleurs, bien que l'article mobilise plusieurs algorithmes de ML notamment, il n'y a pas d'évaluation comparative approfondie entre eux à part leur accuracy. Une étude plus détaillée des performances aurait été utile pour guider le choix des méthodes selon les contextes ou le type de données.

On pourrait aussi dire que l'article se limite à la prévision du nombre de lits, sans considérer d'autres ressources hospitalières comme le personnel soignant ou les équipements, qui sont pourtant indispensables à une planification globale. Les résultats sont également présentés de manière déterministe, sans intervalles de confiance ni scénarios d'incertitude, ce qui limite la robustesse des décisions prises à partir de ces prévisions. Enfin, bien que l'étude identifie que 6 % des patients sont des enfants avec un séjour plus long, elle ne propose aucune action concrète en réponse. Peut-être qu'il aurait été pertinent de recommander, par exemple, la création d'un service pédiatrique dédié.

Enfin, l'approche reste assez centrée sur un seul hôpital et bien qu'elle tienne compte de certains facteurs tels que les fêtes nationales, elle ne tient pas compte de facteurs externes tels que les flux inter-hôpitaux pour aider à gérer la surcharge des patients. Une extension vers des modèles multi-hôpitaux aurait renforcé la portée du travail.

Conclusion

Ce projet nous a permis de découvrir et de mettre en pratique des méthodes de ML appliquées à un problème très concret et critique : la gestion des lits hospitaliers. Dans l'article étudié, nous avons vu comment l'analyse de données, la classification et la prévision peuvent être combinées pour anticiper les besoins en ressources dans un hôpital.

Notre réimplémentation, même simplifiée, montre que le ML peut produire des résultats exploitables à partir de données hospitalières réelles. En ajoutant une couche d'optimisation, nous avons également souligné que la PM garde toute sa pertinence, non pas en opposition au ML, mais en complément, pour intégrer des contraintes concrètes et proposer des décisions réalistes via l'alliance des deux approches.