

# NO-SHOW : prédire grâce au machine-learning

Projet KADUCEO



Salomé CHEVAILLER - Camille MARC - Thibault TEISSEIRE



# SOMMAIRE

## 01. INTRODUCTION

Présentation du commanditaire, des enjeux du projet et du sujet

## 02. RÉALISATIONS

Recherche d'articles, Machine Learning, Maquettes

## 03. GESTION DE PROJET

Organisation de l'équipe et outils utilisés

## 04. CONCLUSION

Retour sur le projet, pistes d'amélioration, ouverture

# 01.

## INTRODUCTION

Présentation du  
commanditaire, des enjeux  
du projet et du sujet



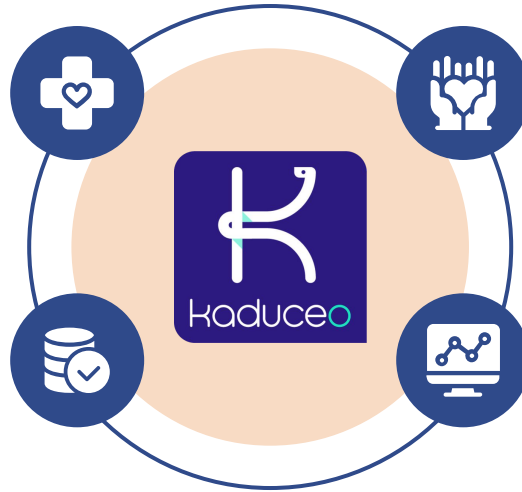
# LE COMMANDITAIRE

## Start-up toulousaine

Déploie des solutions logicielles basées sur l'**analyse** de **données**

## Jean-Baptiste EXCOFFIER

Ingénieur Data Scientist depuis 2019



## Fondée en 2014

Pour accompagner les **professionnels** de **santé** dans l'amélioration de leurs pratiques

## En 2022 ?

Solution no-code d'**aide à la décision**

# QU'EST-CE QUE LE NO-SHOW?

- Patient qui ne se présente **PAS** à son rdv
- Sans l'annuler et sans avertir le personnel

42% des patients <sup>[1]</sup>

Ne se présenteraient pas à leur rendez-vous aux Etats-Unis



# LES RAISONS ?

**Argent**

49 millions de patients <sup>[1]</sup>

**Oubli**

48% des patients <sup>[2]</sup>

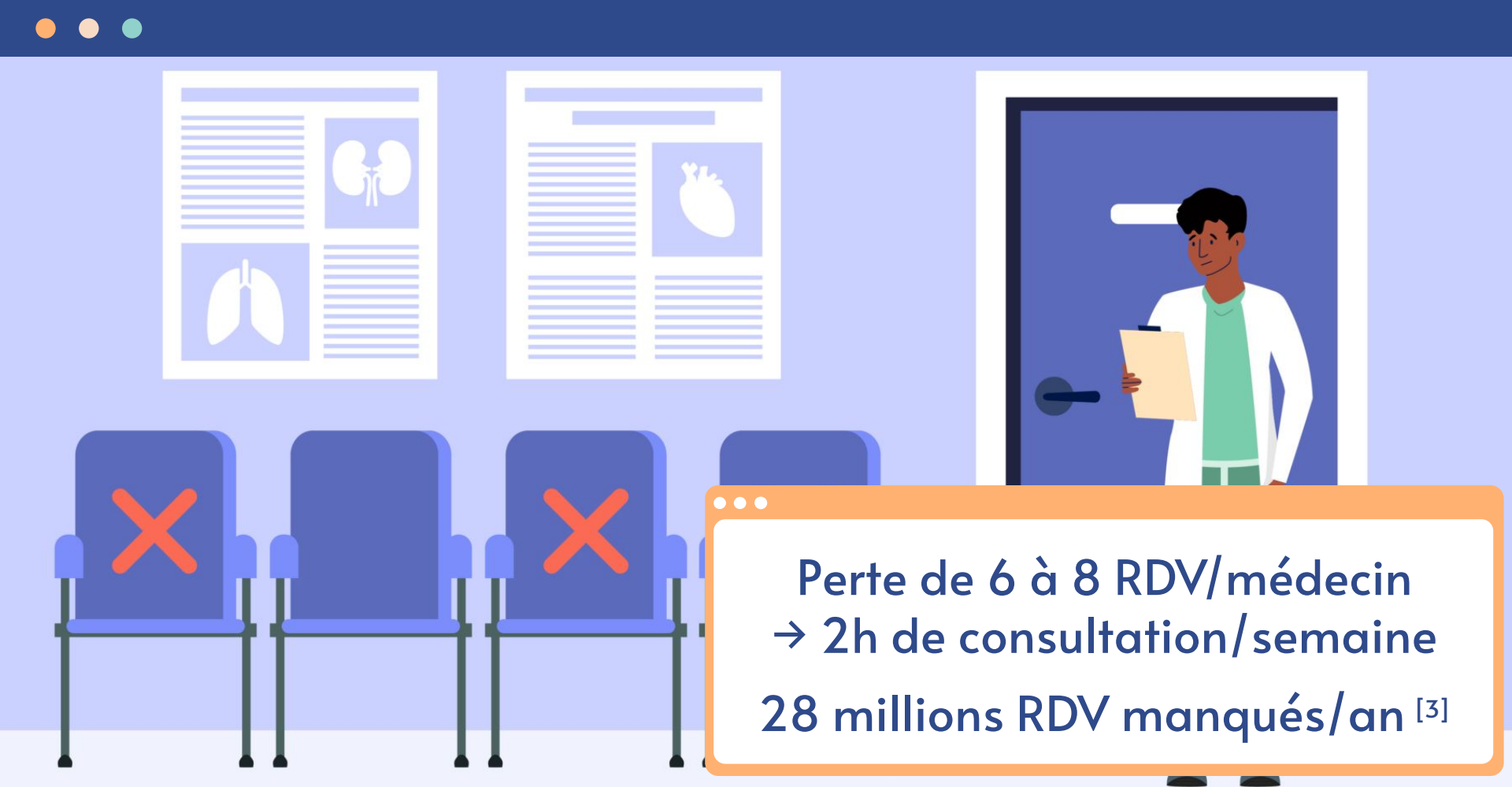
**Transports**

Manque d'autonomie

**Peur,  
démographie...**

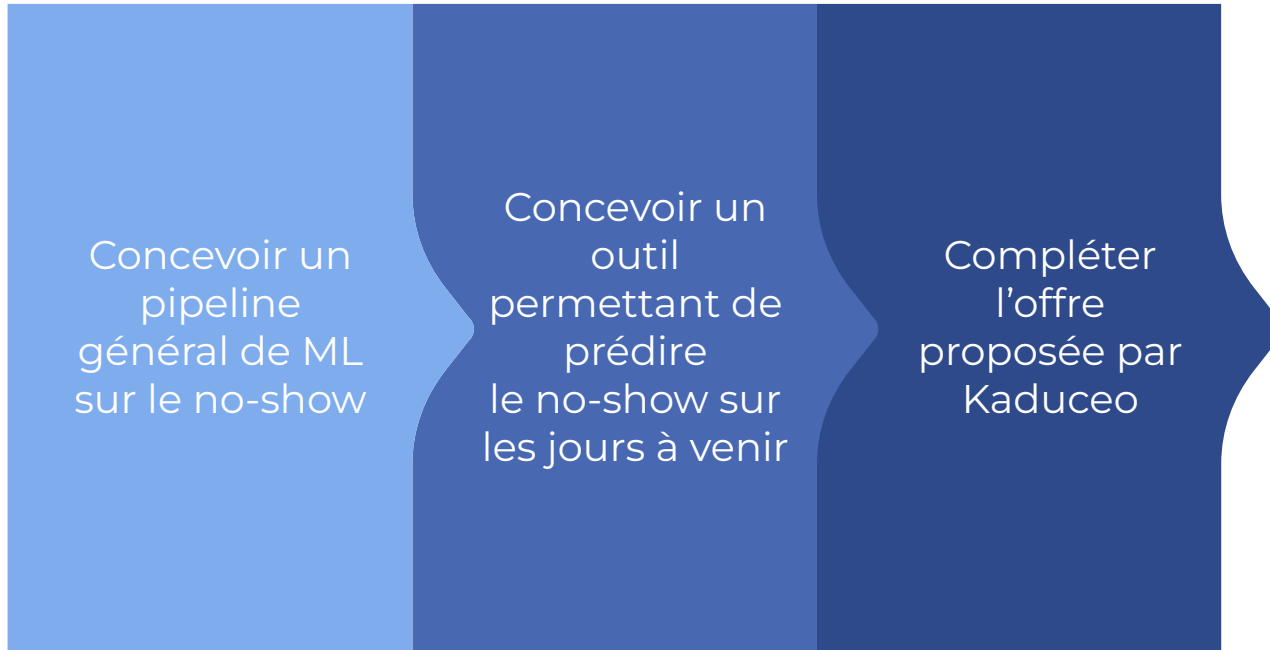
[1] Tšernov, K. (2022, 12 janvier) The Effect of No-Show Appointments on Patients and Hospitals

[2] Tranthimy, L. (2019, 11 avril) Rendez-vous non honorés : les « lapins » se banalisent et pourrissent l'agenda médical



Perte de 6 à 8 RDV/médecin  
→ 2h de consultation/semaine  
28 millions RDV manqués/an <sup>[3]</sup>

# ENJEUX ET OBJECTIFS





# 02.

## RÉALISATIONS

Recherche d'articles,  
Machine Learning,  
Maquettes



# RÉALISATIONS



## Recherche d'articles

Relatifs au no-show afin d'en extraire les variables explicatives de ce comportement



## Machine Learning

Manipulation des variables et d'algorithmes



## Maquettes

À destination des professionnels pour présenter les résultats des algorithmes

# RÉALISATIONS



## Recherche d'articles

Relatifs au no-show afin d'en extraire les variables explicatives de ce comportement



## Machine Learning

Manipulation des variables et d'algorithmes



## Maquettes

À destination des professionnels pour présenter les résultats des algorithmes

# VARIABLES EXPLICATIVES DE LA LITTÉRATURE

Catégories	Données démographiques du patient	Comportement du patient	Rendez-vous
Variables	Age (100%) [7][8][9][10][11][12][13][14]	Taux d'absence (37,5%) [10][12][14]	Jour de la semaine (62,5%) [9][10][11][12][14]
	Genre (75%) [7][9][11][12][13][14]	Absence au RDV précédent (25%) [8][11]	Moment de la journée (75%) [9][10][11][12][13][14]
	Langue parlée (37,5%) [7][11][14]	Délai entre la prise de RDV et le RDV (62,5%) [8][9][11][13][14]	Mois (50%) [8][9][11][14]
	Ethnie (50%) [7][11][12][14]	Nombre de jours depuis le dernier RDV (25%) [10][11]	Saison (25%) [10][11]
	Province d'origine (12,5%) [9]	Statut de la dernière visite (12,5%) [11]	Météo (12,5%) [11]
	Code postal (12,5%) [11]	Expérience précédente (12,5%)	Jour de vacances (12,5%) [11]
	Distance du lieu de RDV (62,5%) [8][9][10][11][14]	Annulation du RDV précédent (12,5%) [11]	Type de l'examen (12,5%) [8]
	Région géographique (12,5%) [12]	Mode de prise du RDV (12,5%) [11]	Spécialité (37,5%) [8][11][12]
	Statut économique (12,5%) [11]	Satisfaction (12,5%) [11]	Statut du patient (37,5%) [9][12][14]
	Statut marital (25%) [10][11]	RDV pris 14 jours avant (12,5%) [10]	SMS de confirmation (25%) [8][13]
	Niveau d'éducation (12,5%) [11]	Heure de la prise du RDV (12,5%) [9]	Lieu du RDV (37,5%) [9][11][12]
	Emploi (12,5%) [11]		
	Accès au téléphone (12,5%) [11]		
	Religion (25%) [11][14]		
	Nombre de personnes dans le foyer (12,5%) [14]		
Catégories	Docteur	Assurance maladie	Caractéristiques cliniques
Variables	Note en ligne (12,5%) [9]	% des coûts couverts (12,5%) [10]	Diagnostic (12,5%) [11]
	Type (12,5%) [9]	En possession (50%) [7][10][11][14]	Durée du suivi (12,5%) [11]
	Titre (12,5%) [9]	Titulaire de l'assurance (12,5%) [14]	Nombre d'admissions à l'hôpital (12,5%) [10]
			Nombre de précédents RDV (25%) [10][11]
			Hypertension (12,5%) [13]
			Handicap (12,5%) [13]
			Index Charlson (12,5%) [10]
			Diabète (25%) [10][13]
			Maladie cardiaque (12,5%) [10]
			Dépression majeure/profonde (12,5%) [10]
			AVC ou démence (12,5%) [10]
			Douleur chronique (12,5%) [10]
			Insuffisance cardiaque (12,5%) [10]
			Maladie pulmonaire chronique (12,5%) [10]
			Dépendance aux drogues (12,5%) [10]
			Consommation de stupéfiants (12,5%) [10]
Légende	50% d'apparition		
	> 50% d'apparition		

# VARIABLES EXPLICATIVES DE LA LITTÉRATURE

Catégories	Données démographiques du patient	Comportement du patient	Rendez-vous
Variables	Age (100%) [7][8][9][10][11][12][13][14]	Taux d'absence (37,5%) [10][12][14]	Jour de la semaine (62,5%) [9][10][11][12][14]
	Genre (75%) [7][9][11][12][13][14]	Absence au RDV précédent (25%) [8][11]	Moment de la journée (75%) [9][10][11][12][13][14]
	Langue parlée (37,5%) [7][11][14]	Délai entre la prise de RDV et le RDV (62,5%) [8][9][11][13][14]	Mois (50%) [8][9][11][14]
	Ethnie (50%) [7][11][12][14]	Nombre de jours depuis le dernier RDV (25%) [10][11]	Saison (25%) [10][11]
	Province d'origine (12,5%) [9]	Statut de la dernière visite (12,5%) [11]	Météo (12,5%) [11]
	Code postal (12,5%) [11]	Expérience précédente (12,5%)	Jour de vacances (12,5%) [11]
	Distance du lieu de RDV (62,5%) [8][9][10][11][14]	Annulation du RDV précédent (12,5%) [11]	Type de l'examen (12,5%) [8]
	Région géographique (12,5%) [12]	Mode de prise du RDV (12,5%) [11]	Spécialité (37,5%) [8][11][12]
	Statut économique (12,5%) [11]	Satisfaction (12,5%) [11]	Statut du patient (37,5%) [9][12][14]
	Statut marital (25%) [10][11]	RDV pris 14 jours avant (12,5%) [10]	SMS de confirmation (25%) [8][13]
	Niveau d'éducation (12,5%) [11]	Heure de la prise du RDV (12,5%) [9]	Lieu du RDV (37,5%) [9][11][12]
	Emploi (12,5%) [11]		
	Accès au téléphone (12,5%) [11]		
	Religion (25%) [11][14]		
	Nombre de personnes dans le foyer (12,5%) [14]		
Catégories	Docteur	Assurance maladie	Caractéristiques cliniques
Variables	Note en ligne (12,5%) [9]	% des coûts couverts (12,5%) [10]	Diagnostic (12,5%) [11]
	Type (12,5%) [9]	En possession (50%) [7][10][11][14]	Durée du suivi (12,5%) [11]
	Titre (12,5%) [9]	Titulaire de l'assurance (12,5%) [14]	Nombre d'admissions à l'hôpital (12,5%) [10]
			Nombre de précédents RDV (25%) [10][11]
			Hypertension (12,5%) [13]
			Handicap (12,5%) [13]
			Index Charlson (12,5%) [10]
			Diabète (25%) [10][13]
			Maladie cardiaque (12,5%) [10]
			Dépression majeure/profonde (12,5%) [10]
			AVC ou démence (12,5%) [10]
			Douleur chronique (12,5%) [10]
			Insuffisance cardiaque (12,5%) [10]
			Maladie pulmonaire chronique (12,5%) [10]
			Dépendance aux drogues (12,5%) [10]
			Consommation de stupéfiants (12,5%) [10]
Légende	50% d'apparition		
	> 50% d'apparition		

Taux d'apparition



Nombre total  
d'apparitions

Nombre total de  
documents

# VARIABLES EXPLICATIVES DE KADUCEO

Catégories	Données démographiques du patient	Comportement du patient	Rendez-vous	Assurance maladie	Caractéristiques cliniques
Variables	Age	Venue lors des 3 derniers jours	Humidité	Type d'assurance	Durée de séjour
	Sexe	Venue lors des 7 derniers jours	Jour de la semaine		Hospitalisation lors des 3 derniers jours
		Venue lors du mois précédent	Pluie		Hospitalisation lors des 7 derniers jours
		Nombre de venues les 3 derniers mois	Pression au niveau de la mer		Hospitalisation lors du dernier mois
		Consultation lors des 3 derniers jours	Vitesse du vent		
		Consultation lors des 7 derniers jours	Température		
		Consultation lors du mois précédent			
		Taux d'absentéisme passé			
		Statut de la précédente visite			

# VARIABLES EXPLICATIVES DE KADUCEO

Catégories	Données démographiques du patient	Comportement du patient	Rendez-vous	Assurance maladie	Caractéristiques cliniques
Variables	Age	Venue lors des 3 derniers jours	Humidité	Type d'assurance	Durée de séjour
	Sexe	Venue lors des 7 derniers jours	Jour de la semaine		Hospitalisation lors des 3 derniers jours
		Venue lors du mois précédent	Pluie		Hospitalisation lors des 7 derniers jours
		Nombre de venues les 3 derniers mois	Pression au niveau de la mer		Hospitalisation lors du dernier mois
		Consultation lors des 3 derniers jours	Vitesse du vent		
		Consultation lors des 7 derniers jours	Température		
		Consultation lors du mois précédent			
		Taux d'absentéisme passé			
		Statut de la précédente visite			

# VARIABLES EXPLICATIVES À AJOUTER

Catégories	Données démographiques du patient	Comportement du patient	Rendez-vous	Caractéristiques cliniques
Variables	Distance du lieu de RDV	Absence à un RDV précédent	Moment de la journée	Hypertension
		Délai entre la prise de RDV et le RDV	Mois	Diabète
		Expérience précédente	Spécialité	
		Heure de la prise de RDV	SMS de confirmation	
		Statut du patient		

# RÉALISATIONS



## Recherche d'articles

Relatifs au no-show afin  
d'en extraire les variables  
explicatives de ce  
comportement



## Machine Learning

Manipulation des  
variables et  
d'algorithmes



## Maquettes

À destination des  
professionnels pour  
présenter les résultats  
des algorithmes



# JEU DE DONNÉES : CARACTÉRISTIQUES



# JEU DE DONNÉES : REMARQUES

- Des **résultats contraires** à ceux attendus (rappel SMS)
- Une période de récolte des données trop courte : **moins de 2 mois**
- **Manque de données** cruciales : spécialité du RDV, distance domicile-lieu de RDV, etc.

# MANIPULATION DU JEU DE DONNÉES



# MANIPULATION DU JEU DE DONNÉES

Modèles	Score F1	Accuracy	
KNeighbors	0.51	0.8402	
Naive Bayes	0.39	0.4620	<b>Légende :</b>
Decision tree	0.54	0.8116	Meilleur score
Random forest	0.53	0.8517	Deuxième
XGboost	0.50	0.8541	Troisième
AdaBoost	0.50	0.8523	
LinearSVC	0.49	0.8485	

Résultats des différents modèles

# RÉALISATIONS



## Recherche d'articles

Relatifs au no-show afin d'en extraire les variables explicatives de ce comportement



## Machine Learning

Manipulation des variables et d'algorithmes

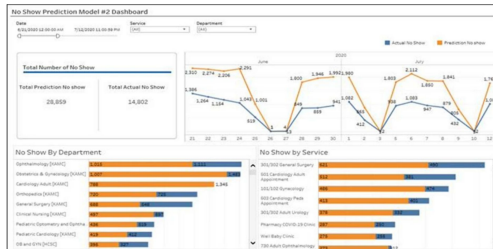


## Maquettes

À destination des professionnels pour présenter les résultats des algorithmes

# INTERPRÉTATION DES MAQUETTES

- PAS de **sur-staffing**
  - Prévoir 1 médecin au lieu de 3
- **Surbooking**
  - Prévoir plus de patients que de créneaux



Daghistani, T., AlGhamdi, H., Alshammari, R., & AlHazme, R. H. (2020). Predictors of outpatients' no-show: big data analytics using apache spark. In Journal of Big Data (Vol. 7, Issue 1). Springer Science and Business Media LLC.

Filtrer par : Période ▼ Pôle ▼ Service ▼ Professionnel ▼

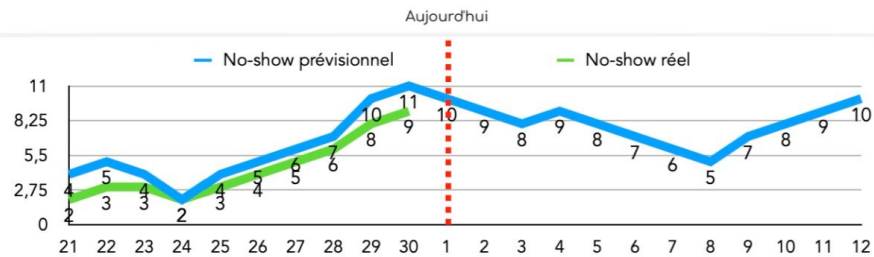
### Nombre total de no-shows

No-show prévisionnel

366

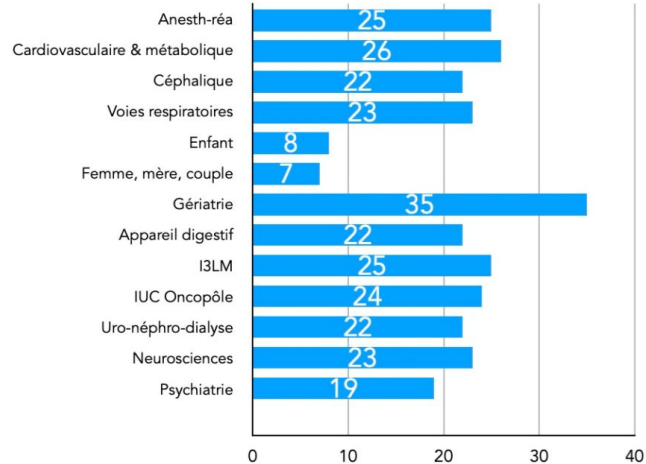
No-show réel

293



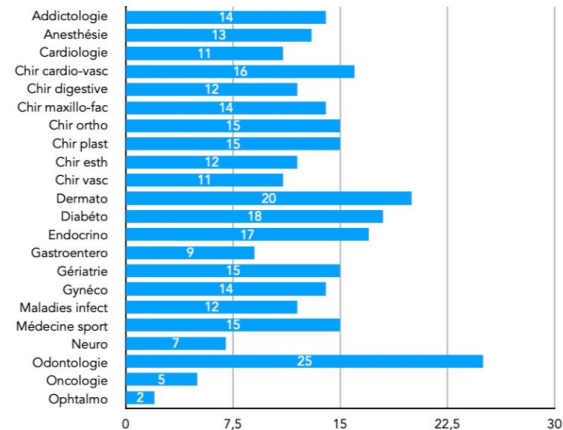
### No-show par pôle

No-show prévisionnel



### No-show par service

No-show prévisionnel



Filtrer par : Période ▼ Pôle ▼ Service ▼ Professionnel ▼

Professionnel ×

Dr GOUBIN ×

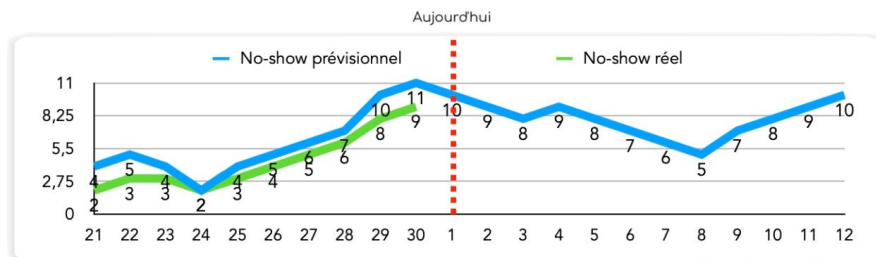
### Nombre total de no-shows

No-show prévisionnel

366

No-show réel

293



### Liste des patients

Identification visite	Identification patient	Prédiction	Raison
0	M. Tom BOLLA	0.1%	Pluie, 8 no-shows précédents
1	Mme. Dina MYTHE	4.2%	Pluie
2	M. Vincent TIME	0.4%	1 no-show précédent
3	Mme. Félicie TATION	36.6%	18 ans
0	M. Paul HEMIQUE	91.2%	Distance domicile > 50 km
1	M. Eddy MOITOU	7.6%	2 no-shows précédents



# 03.

## GESTION DE PROJET

Organisation de l'équipe et  
outils utilisés



# GESTION DE L'ÉQUIPE

Salomé CHEVAILLER



Camille MARC



Thibault TEISSEIRE



*Chef de projet*

# GESTION DU PROJET

▼ Tableau Sprint 1 10 oct. – 14 oct. (6 tickets)

0 0 0

Démarrer un sprint



- |   |           |  |
|---|-----------|--|
| <input checked="" type="checkbox"/> KADUCEO-1 S'informer sur la plateforme de Machine Learning à destination des hôpitaux : <a href="https://kaduceo.com/no-code-machine-learning-don...">https://kaduceo.com/no-code-machine-learning-don...</a> | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-2 Consulter les onglets Ressources et R&D du site   | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-3 Lire les 5 articles scientifiques pour compréhension générale de la problématique du no-show  | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-4 Rechercher d'autres articles scientifiques connexes aux 5   | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-5 Lister les variables explicatives référencées dans la littérature   | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-6 Tester la plateforme Kaducéo  | À FAIRE ▼ |  |

+ Créer un ticket

# GESTION DU PROJET

▼ Tableau Sprint 1 10 oct. – 14 oct. (6 tickets)

0 0 0

Démarrer un sprint



- |   |           |  |
|---|-----------|--|
| <input checked="" type="checkbox"/> KADUCEO-1 S'informer sur la plateforme de Machine Learning à destination des hôpitaux : <a href="https://kaduceo.com/no-code-machine-learning-don...">https://kaduceo.com/no-code-machine-learning-don...</a> | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-2 Consulter les onglets Ressources et R&D du site   | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-3 Lire les 5 articles scientifiques pour compréhension générale de la problématique du no-show  | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-4 Rechercher d'autres articles scientifiques connexes aux 5   | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-5 Lister les variables explicatives référencées dans la littérature   | À FAIRE ▼ |  |
| <input checked="" type="checkbox"/> KADUCEO-6 Tester la plateforme Kaducéo  | À FAIRE ▼ |  |

+ Créer un ticket

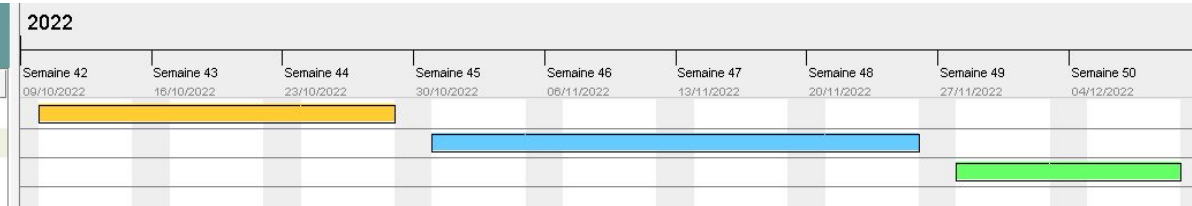


&



# GESTION DU PROJET

Annuler indentation		
project		
Nom	Date de début	Date de fin
• Recherche de variables explicatives	10/10/2022	28/10/2022
• Manipulation du jeu de données	31/10/2022	25/11/2022
• Elaboration de maquettes	28/11/2022	09/12/2022

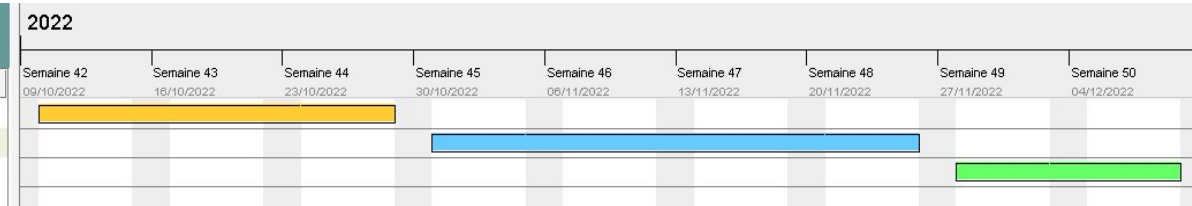


VS

# GESTION DU PROJET

Annuler indentation project

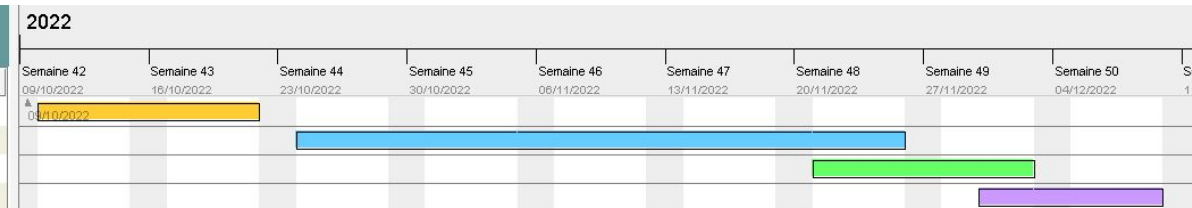
Nom	Date de début	Date de fin
• Recherche de variables explicatives	10/10/2022	28/10/2022
• Manipulation du jeu de données	31/10/2022	25/11/2022
• Elaboration de maquettes	28/11/2022	09/12/2022



VS

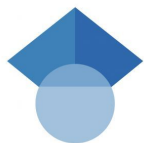
GANTT project

Nom	Date de début	Date de fin
• Recherche de variables explicatives	10/10/2022	21/10/2022
• Manipulation du jeu de données	24/10/2022	25/11/2022
• Elaboration de maquettes	21/11/2022	02/12/2022
• Recherche de nouveaux cas d'usage	30/11/2022	09/12/2022



# OUTILS UTILISÉS

## Outils de recherche



Google Scholar



**CONNECTED  
PAPERS**

## Outils techniques

# OUTILS UTILISÉS

## Outils de recherche



## Outils techniques



Google Slides





# 04.

## CONCLUSION

Retour sur le projet, pistes  
d'amélioration, ouverture



# CONCLUSION & AMÉLIORATIONS



## Objectifs atteints !

Etude de la littérature,  
modèles prédictifs,  
maquettes



## Projet très formateur

Recherche  
bibliographique  
approfondie et gestion  
de projet



## Amélioration ML

Performance des  
modèles, courbe ROC,  
aire AUC



## Amélioration maquettes

Utilisation Figma aurait  
été intéressante et rendu  
la maquette interactive

# Prochaine étape ? L'étude du Sepsis !

- Identification précoce des patients atteints de sepsis
- Sepsis : état aigu de dysrégulation dû à une infection, entraînant la perte de fonction des organes et un risque vital pour le patient

## Raisons :

1. Mal appréhendé par professionnels de santé
2. 57 000 décès / an en France
3. 16 000 € / hospitalisation

Kong, G., Lin, K., & Hu, Y. (2020). Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. In *BMC Medical Informatics and Decision Making* (Vol. 20, Issue 1). Springer Science and Business Media LLC.

Kong et al. *BMC Medical Informatics and Decision Making* (2020) 20:251  
<https://doi.org/10.1186/s12911-020-01271-2>

BMC Medical Informatics and  
Decision Making

## RESEARCH ARTICLE

Open Access

## Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU



Guilan Kong<sup>1,2\*</sup>, Ke Lin<sup>1,2</sup> and Yonghua Hu<sup>3,4</sup>

### Abstract

**Background:** Early and accurate identification of sepsis patients with high risk of in-hospital death can help physicians in intensive care units (ICUs) make optimal clinical decisions. This study aimed to develop machine learning-based tools to predict the risk of hospital death of patients with sepsis in ICUs.

**Methods:** The source database used for model development and validation is the medical information mart for intensive care (MIMIC) III. We identified adult sepsis patients using the new sepsis definition Sepsis-3. A total of 86

# BIBLIOGRAPHIE

- [1] Tšernov, K. (2022, 12 janvier) The Effect of No-Show Appointments on Patients and Hospitals  
<https://www.qminder.com/blog/queue-management/no-shows-affect-hospitals/>
- [2] Tranthimy, L. (2019, 11 avril) Rendez-vous non honorés : les « lapins » se banalisent et pourrissent l'agenda médical (et les patients le savent !)  
<https://www.lequotidiendumedecin.fr/archives/rendez-vous-non-honores-les-lapins-se-banalissent-et-pourrissent-lagenda-medical-et-les-patients-le>
- [3] (2017, 17 février) Doctors want pay for 'no-shows'  
<https://www.connexionfrance.com/article/Archive/Doctors-want-pay-for-no-shows>
- Daghistani, T., AlGhamdi, H., Alshammari, R., & AlHazme, R. H. (2020). Predictors of outpatients' no-show: big data analytics using apache spark. In Journal of Big Data (Vol. 7, Issue 1). Springer Science and Business Media LLC.  
<https://doi.org/10.1186/s40537-020-00384-9>
- Kong, G., Lin, K., & Hu, Y. (2020). Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. In BMC Medical Informatics and Decision Making (Vol. 20, Issue 1). Springer Science and Business Media LLC.  
<https://doi.org/10.1186/s12911-020-01271-2>

Merci de  
votre  
attention !

