

# Application de prédiction de la dose du médicament Tacrolimus chez un transplanté rénal

MEDICA∞M

Encadrants:

Académique : Dr. Raoudha CHEBIL

De l'entreprise: Mr. Majdi Karray

Réalisé par:

Soud Mohamed

Kammoun Mohamed Amine

# Plan

- Introduction
- Élaboration du modèle de Machine learning
- Analyse et spécification des besoins
- Conception
- Réalisation
- Conclusion et perspective

# 1. Introduction

1.1 Contexte professionnel et académique

1.2 Problématique

1.3 Critique de l'existant

1.4 Solution proposée

## Contexte professionnel et académique



X

MEDICA<sup>+</sup>COM

## Problématique

La greffe rénale est une procédure chirurgicale à haut risque et ce risque reste important même dans la période post-opératoire vu que l'immunité du patient est faible.

Dans cette phase, l'administration précise du médicament Tacrolimus est essentielle, car il possède une marge thérapeutique étroite.

Un dosage précis est nécessaire afin de minimiser le risque de rejet du rein.

Cependant, la dose prescrite par les médecins actuellement ne prend pas en considération tous les facteurs et le taux d'erreur est important

## Solution proposée

Conception et développement d'une application Web dotée d'un modèle de machine learning visant à :

- prédire la dose optimale du médicament Tacrolimus.
- la digitalisation du suivi du traitement, permettant ainsi une surveillance plus précise et personnalisée du patient après la greffe rénale.

# 2. Machine Learning

2.1 Étude théorique

2.2 Méthodologie Adoptée

2.3 Préparation des données

2.4 Implémentation du modèle

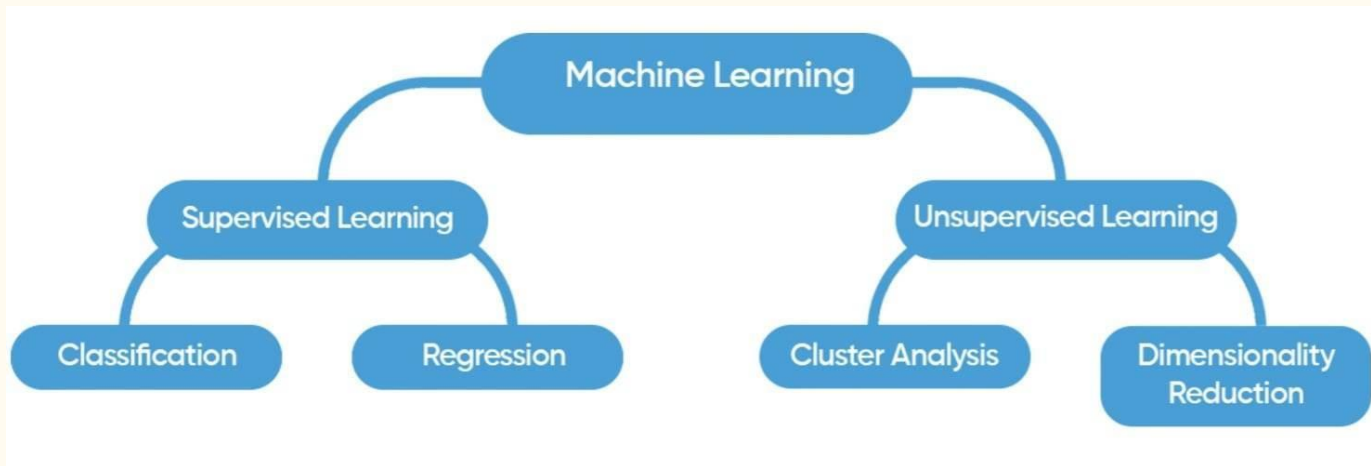
2.5 Évaluation du modèle

## 2.1 Étude théorique



# Type d'apprentissage

- Ce projet s'inscrit dans le cadre de **l'apprentissage supervisé**.
- Comme nous prévoyons une valeur continue, nous utiliserons **la régression**.



# Algorithmes candidats

- **Lasso Regression**
- **Ridge Regression**
- **ElasticNet Regression**
- **SDG (Stochastic Gradient Descent)**
- **RandomForests**
- **XGBoost**
- **Neural Networks**

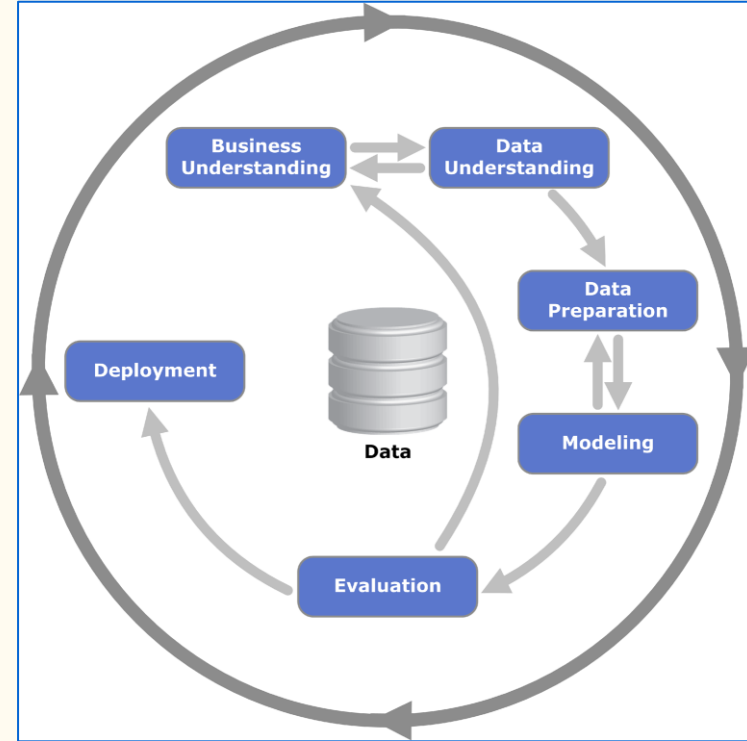
	Lasso	Ridge	ElasticNet	SDG	Random Forests	XGBoost	Neural Network
Problem Type	Régression	Régression	Régression	Both	Both	Both	Both
Results interpretable by user?	Yes	Yes	Yes	Yes	A little	A little	No
Easy to explain algorithm?	Somewhat	Yes	Yes	Yes	No	No	No
Average predictive accuracy	Lower	Lower	Lower	Moderate to high	Higher	Higher	Higher
Training speed	Fast	Fast	Moderate	Fast	Slow	Slow	Slow
Prediction speed	Fast	Fast	Fast	Fast	Moderate	Fast	Fast
Amount of parameter tuning needed	Minimal	Minimal	Minimal	Some	Some	Some	Lots
perform well with a small number of observations?	Yes	Yes	Yes	Yes	Yes	Yes	No
Handles lots of irrelevant features well?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Automatically learns feature interactions?	No	No	No	No	Yes	Yes	Yes

## 2.2 Méthodologie Adoptée

# CRISP-DM

(Cross Industry Standard Process for Data Mining)

- C'est une méthodologie adoptée pour la réalisation du modèle.
- Le cycle de vie de cette méthodologie est illustré dans cette figure.



## 2.3 Préparation des données

# Collecte de données

La collecte de données n'était pas une mission de ce projet.

Des enregistrements de données de test proviennent de différents patients ayant subi une transplantation rénale en Tunisie depuis 1991, et d'autres données ont été ajoutées à partir de différents pays.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
Gender	Age	Height (cm)	Weight (kg)	Living donor	Hemoglobin	Leukocyte	Serum creatinine	Total bilirubin	Albumin	Hypertension	Diabetes	Anemia	Cardiac insufficiency	Calcium channel blocker	Metoprolol	Omeprazole	Furosemide	ACEI /ARA*	Cephalosporin	Infected	Initial_dose	Stable_dose	Dose_group	CYP3A5/698AG	ABCB1/3435CT	ABCB1/129TC	ABCB1/1236CT	ABCB1/2677GT
male	25	NA	67.5	1	NA	NA	NA	NA	NA	NA	NA	1	NA	1	1	1	1	0	1	1	6	0.5	1	GG	NA	TT	CT	GT
male	24	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	1	NA	0	1	1	1	0	1	1	1	0.5	1	GG	NA	NA	NA	NA
male	20	NA	NA	1	NA	NA	NA	NA	NA	1	NA	1	NA	1	1	0	0	1	1	0	8	0.5	1	NA	NA	NA	NA	NA
female	37	163	52	0	108	4.16	708	7.8	43.5	1	0	1	0	1	0	0	0	0	1	1	6	1	1	GG	CC	CC/CT	CT	GT
male	48	NA	56	1	NA	NA	NA	NA	NA	NA	NA	1	NA	1	0	1	1	0	1	1	2	1	1	GG	CC	TT	CC	GT
male	41	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	0	0	1	0	1	0	7	1	1	NA	CT	NA	CT	GT
male	37	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1	1	1	0	0	0	6	1	1	GG	CT	TT	TT	GT
male	37	168	65	1	106	10.83	142	7.9	42.2	1	0	0	0	1	0	1	0	0	1	0	4	1	1	GG	CT	TT	TT	GT
male	36	178	52	0	130	11.3	134	22.8	NA	0	0	NA	NA	0	0	NA	0	0	NA	NA	4	1	1	GG	CT	TT	TT	GT
male	44	173	NA	1	122	9.3	157	7.7	NA	0	0	NA	NA	0	1	NA	0	1	NA	NA	2	1	1	GG	CC	TT	CC	GG
female	40	153	43	0	146	10.99	818	7.4	45.3	1	0	0	0	1	1	0	1	0	1	0	6	1	1	GG	CT	TT	CT	GT
male	49	165	51	0	125	9.08	117	NA	NA	1	0	0	0	0	0	0	0	0	0	0	6	1	1	GG	CT	CC/CT	CT	TT
male	35	175	51	1	140	4.9	107	23	NA	0	0	NA	NA	0	0	NA	0	0	NA	NA	2	1	1	GG	CC	TT	CT	GG
male	36	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	1	NA	1	0	1	1	0	1	0	7	1	1	NA	CT	TT	NA	GT
male	32	160	55	0	154	8.1	111	19.4	49.8	1	0	1	0	1	1	1	1	0	1	0	2	1	1	GG	CT	TT	CC	GG
male	18	159	57	1	102	13.2	75	14.5	44.9	1	0	1	0	1	1	1	0	0	1	1	5	1	1	GG	CC	TT	CT	GG
male	27	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	1	NA	0	0	0	1	0	1	0	8	1	1	GG	TT	TT	TT	NA
male	50	172	73	1	92	5.5	134	4.5	NA	0	0	NA	NA	0	0	NA	0	0	NA	NA	2	1	1	GG	CT	TT	CT	TT
male	33	175	51.8	1	99	8.1	816	4.2	41.8	1	0	0	0	1	0	1	1	0	1	0	6	1	1	GG	CC	TT	CT	TT
male	45	165	57	1	68	3.71	1588	7.1	37.5	1	0	0	0	1	0	0	0	0	1	0	2	1	1	GG	CC	TT	CC	GG
female	36	NA	NA	1	156	8.68	598	16.5	43.9	1	0	1	0	0	0	0	1	0	0	1	4	1	1	GG	CT	TT	CT	TT
female	21	162	51	1	135	8.78	88	5.3	47.4	1	0	0	0	1	0	0	0	0	1	1	4	1	1	GG	CC	TT	TT	GG
male	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	1	NA	0	0	0	0	0	0	0	6	1	1	GG	TT	TT	CT	GT
male	27	NA	47.5	1	NA	NA	NA	NA	NA	NA	NA	1	NA	1	1	1	0	0	1	1	10	1	1	GG	CT	TT	CT	GT
female	41	NA	60	0	NA	NA	NA	NA	NA	NA	NA	1	NA	0	0	1	1	1	1	0	9	1	1	GG	CT	TT	CT	GT
female	52	150	80	0	138	6.88	114	9.1	50.4	1	0	1	0	1	0	0	1	0	0	0	2	1	1	GG	TT	TT	TT	TT
male	26	170	56	0	147	5.4	105	7.9	NA	0	0	NA	NA	0	0	NA	0	0	NA	NA	2	1.25	1	AG	TT	TT	TT	TT
female	37	160	54	1	NA	NA	NA	NA	NA	0	0	NA	NA	0	0	NA	0	0	NA	NA	NA	1.25	1	AG	TT	TT	TT	TT
male	22	164	56	0	115	4.77	77	6.8	42.4	1	0	0	0	1	1	0	1	0	1	1	5	1.5	1	GG	CT	TT	TT	GT
male	20	NA	60	0	NA	NA	NA	NA	NA	1	NA	NA	NA	1	1	0	1	1	1	0	7	1.5	1	GG	CT	NA	CT	GT
male	33	165	59	0	81	8.07	222	11.5	36.3	0	0	1	0	1	0	0	1	0	0	0	6	1.5	1	GG	CT	TT	TT	GT
female	47	NA	NA	0	NA	NA	NA	NA	NA	NA	NA	1	NA	1	1	1	1	1	1	0	8	1.5	1	AG	CT	TT	CT	TT
male	55	NA	75.5	0	NA	NA	NA	NA	NA	NA	NA	1	NA	1	1	1	1	1	1	0	12	1.5	1	AG	CC	TT	CT	GG
male	??	NA	51	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	0	1	1	0	1	0	10	1.5	1	GG	NA	NA	NA	NA



# Nettoyage des données

- Suppression des doublons
- Détection et traitement des valeurs manquantes
- L'encodage des variables catégorielles

```
print(rein_df.isna().sum().sort_values())
```

Calcium channel blocker	0
ABCB1 129TC	0
ABCB1 3435CT	0
CYP1A5 6986AG	0
Dose_group	0
Stable_dose	0
ACEI /ARA*	0
Furosemide	0
Metoprolol	0
ABCB1 1236CT	0
Diabetes	0
ABCB1 2677GT	0
Age	0
Height (cm)	0
Weight (kg)	0
Hypertension	0
Gender	1
Living donor	2
Initial_dose	29
Serum creatinine	62
Hemoglobin	71
Leukocyte	74
Total bilirubin	83
Anemia	130
Omeprazole	130
Cardiac insufficiency	130
Cephalosporin	130
Infected	130
Albumin	157
dtype: int64	

```
#encoding Labels
```

```
data['Gender'] = data['Gender'].map( {'male':1, 'female':2})  
data['CYP3A5 6986AG']=data['CYP3A5 6986AG'].map({'GG':1, 'AG':2, 'AA':3})  
data['ABCB1 3435CT']=data['ABCB1 3435CT'].map({'CC':1, 'CT':2, 'TT':3})  
data['ABCB1 129TC']=data['ABCB1 129TC'].map({'CC/CT':1, 'TT':2})  
data['ABCB1 1236CT']=data['ABCB1 1236CT'].map({'CC':1, 'CT':2, 'TT':3})  
data['ABCB1 2677GT']=data['ABCB1 2677GT'].map({'GG':1, 'GT':2, 'TT':3})
```

## 2.4 Implémentation du modèle

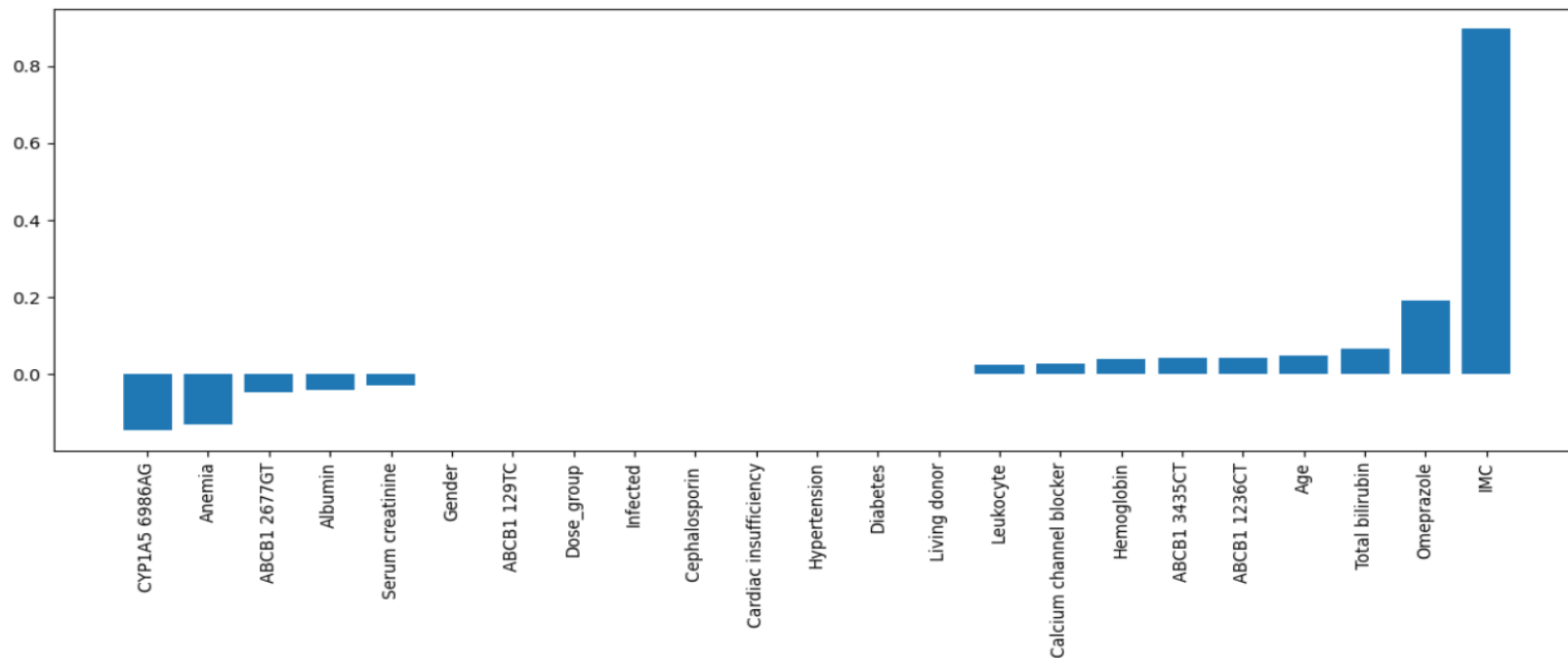
## Étapes d'implémentation

1. Construction du modèle
2. Optimisation et réajustement du modèle

## Construction du modèle

- Technique de train test split
- L'importance des variables
- Instanciation de chaque modèle

# L'importance des variables



## Optimisation et réajustement du modèle

- Cross-Validation
- Hyperparameter Tuning  
(GridSearchCV)



## 2.5 Évaluation du modèle

## Métriques d'évaluation

- MAE (Mean Absolute Error)
- RMSE (Root Mean Square Error)
- R-carré

# Comparaison des résultats

	MAE	RMSE	R-Square	
	Train	Test	Train	Test
XGBoost	0.13323	0.29205	0.903584	0.935189
RF	0.3560215	0.44834	0.669694	0.731750
Lasso	0.474421	0.62519	0.4600706	0.478373
Ridge	0.45688	0.578959	0.50	0.55267
ElasticNet	0.4494	0.58515	0.4945	0.543062
SGD Regressor	0.497	0.6251	0.463247	0.47837

## Choix de l'algorithme

D'après le tableau, on peut constater que XGBoost a fourni les résultats les plus satisfaisants, à la fois sur l'ensemble d'entraînement  $[0,90]$  et sur l'ensemble de test  $[0,93]$ .

# 3. Analyse et spécification des besoins

# Acteurs

- SuperAdmin
- Médecin
- Infirmier

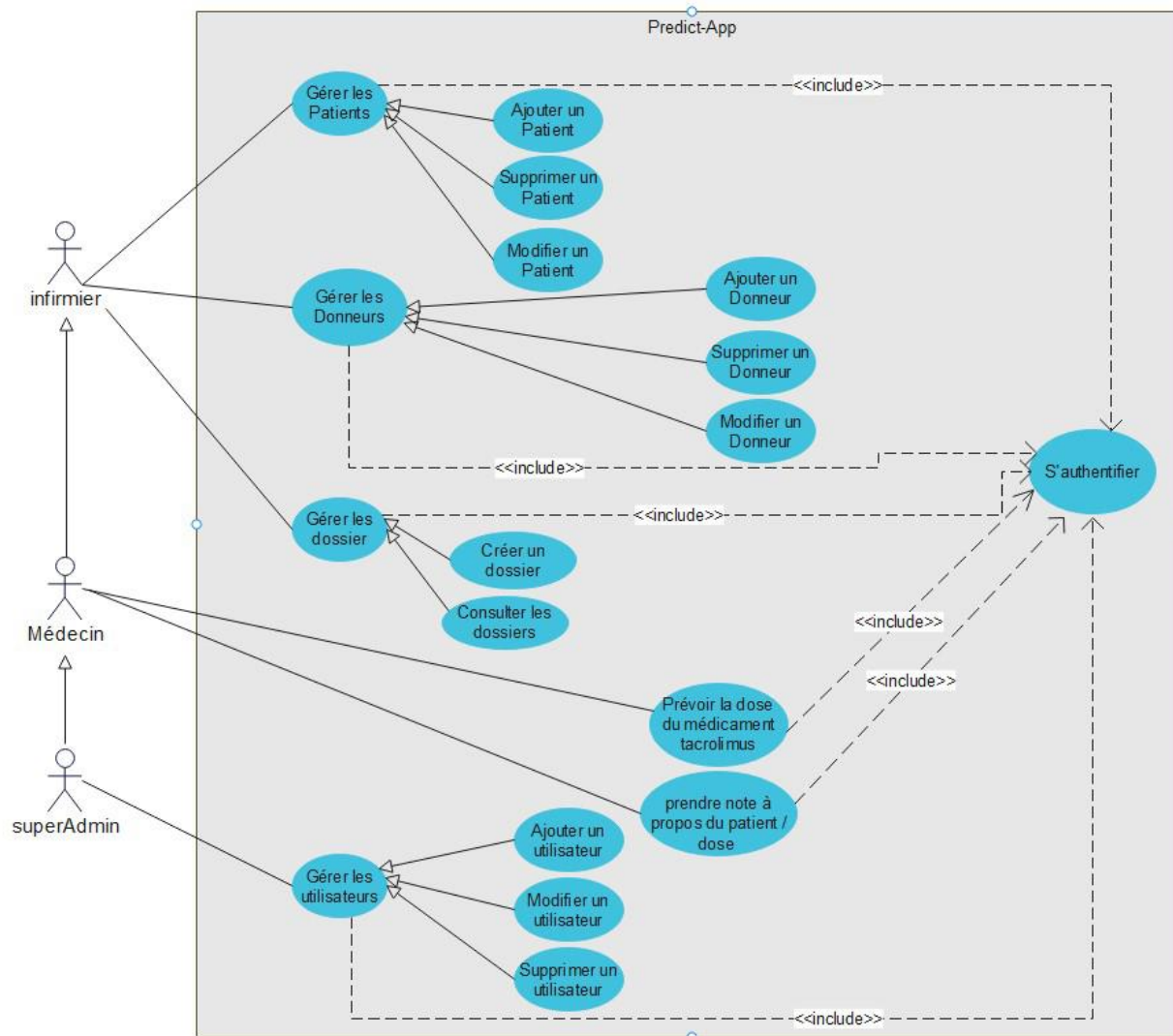


## Besoins fonctionnels

- S'authentifier.
- Gérer les médecins et des infirmiers.(pour le superadmin)
- Gérer les patients, les donneurs et les dossiers.
- Entrer les données d'un patient afin de prévoir la dose optimale pour ce dernier.
- Ajouter des notes/remarques à propos d'un patient bien déterminé.

# Diagramme de cas d'utilisation





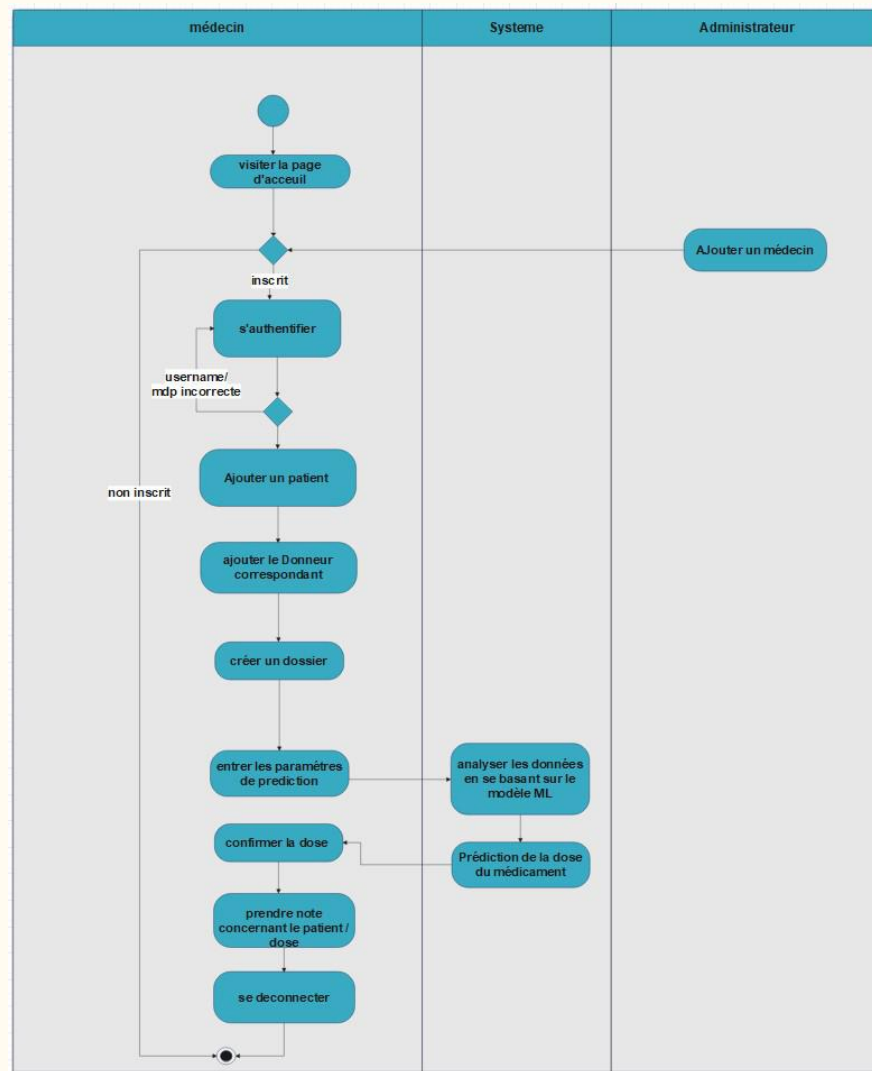
# Besoins non fonctionnels

- L'évolutivité
- La convivialité
- Sécurité



# Diagramme d'activité

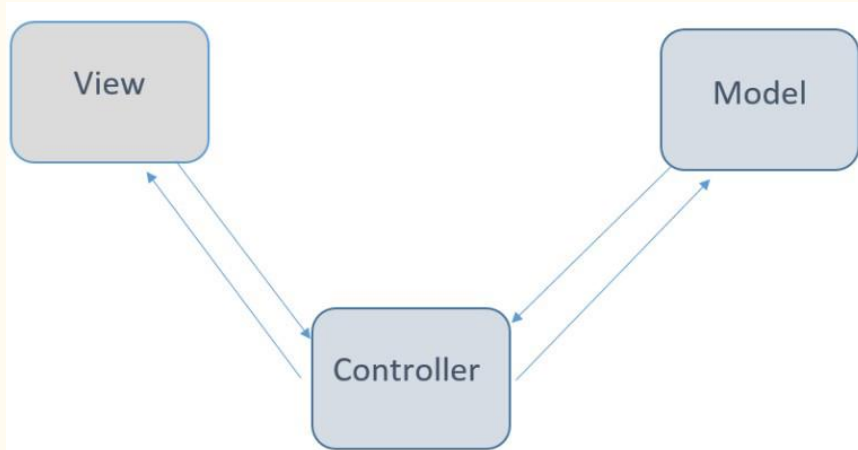
scénario prédiction de la dose du médicament



# 4. Conception

# Architecture logique

# Le modèle architectural MVC



- Model: Permet la gestion des données stockées dans la base de données et leur manipulation.
- View: Permet de présenter les données récupérées du modèle de manière claire et intuitive pour l'utilisateur.
- Controller: Joue le rôle de l'intermédiaire entre le modèle et la vue. Son rôle est de gérer les requêtes de l'utilisateur

# 5. Réalisation

5.1 Méthodologie adoptée

5.2 Technologies

5.3 Démo



# 5.1 Méthodologie adoptée

# Méthodologie Agile Scrum

La méthodologie **Agile Scrum** est une approche de gestion de projet qui fait partie des méthodologies agiles.

- Les fonctionnalités du projet ne sont pas fixes dès le départ.
- Les besoins de l'entreprise évoluent au fil du temps.

## 5.2 Technologies

# Frameworks et bibliothèques



- **Développement web frameworks javascript**

React Js

Bootstrap

node Js

Express JS

- **Système de gestion de base de donnée**

MongoDB

- **Réalisation du modèle bibliothèques python**

scikit-learn

Pandas

## 5.3 Démo

## 6. Conclusion et perspective

# Conclusion

Ce projet met en évidence l'importance de l'intégration de l'apprentissage automatique dans le domaine médical et précisément pour prédire efficacement la dose stable de Tacrolimus.

Les résultats préliminaires encourageants démontrent la faisabilité d'utiliser des approches d'apprentissage automatique pour résoudre des problèmes complexes dans le domaine médical, permettant ainsi une meilleure compréhension des données des patients et des décisions plus éclairées en matière de soins de santé.

## Perspective

Améliorer la précision, enrichir notre base de donnée avec des données réelles.

Généraliser le modèle sur les autres Immunosuppresseurs qui sont aussi des médicaments à marge thérapeutique étroite.



Merci pour votre attention