**Machine Learning Project :**

# **Problem statement and context**

* **Problem Statement :**

The objective of this project is to analyse and predict the most suitable medication for a patient based on specifical attributes. Specifically, we aim to classify wich drug *(DrugA, DrugB, DrugC, DrugX, or DrugY)* should be prescribed given data on:

* Age
* Sex
* Blood Pressure (BP)
* Cholesterol levels
* Sodium-to-Potassium ratio in the blood (Na\_to\_k)

This is a classification problemwhere the target variable is the type of drug (Drug), a categorical feature. The goal is to leverage machine learning to improve the efficiency and accuracy of prescribing drugs based on patient data.

Prédire le type de médicament le plus approprié pour un patient basé sur des caractéristiques telles que l'âge, le sexe, la pression artérielle, le cholestérol et le ratio sodium/potassium dans le sang.

* **Context of the study**

The healthcare industry increasingly leverages machine learning to enhance decision-making in patient care. By analyzing features such as age, gender, blood pressure, cholesterol, and the sodium-to-potassium ratio, machine learning models provide tailored drug prescriptions, improving treatment accuracy and reducing variability【1】【2】. The sodium-to-potassium ratio, a key biomarker for cardiovascular health, is particularly important in drug prescription as it influences hypertension management【3】.

This project has real-world applications, including developing decision-support systems for healthcare providers and integrating machine learning into electronic health records (EHR) for real-time drug recommendations. Furthermore, analyzing patterns in drug prescriptions and patient demographics can inform public health initiatives and policymaking.

# **Exploratory Data Analysis**

L’EDA est importante dans ce projet car c’est une étape qui vise à explorer le dataset et à identifier certains patterns. Par ailleurs, c’est durant cette phase que l’on prépare nos données pour la modélisation.

**Statistiques Descriptives**

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

Figure 1:Descriptive statistics for the variables age and cholesterol level.

D’après (citer une source), le taux de cholesterol (Na\_to\_k) moyen d’une personne se situe autour de [valeur a ; valeur b]. Les relevés de notre echantillon sont autour de 16, ce qui est conforme aux observations mpyennes des études, et les valeurs varient entre 6 et 38. D’après notre jeu de données, l’âge moyen se situe autour de 44 ans, avec des valeurs qui oscillent entre 15 et 74 ans.

Une image contenant diagramme, texte, capture d’écran, Tracé

Description générée automatiquement

**References:**

1. Sutton, R. T., et al. (2020). Machine Learning in Clinical Decision Support: Use Cases, Challenges, and Solutions. *BioMed Research International*. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/32089925/>
2. Bini, S. A. (2018). The Role of Artificial Intelligence in Medicine: Applications and Challenges. *The Journal of Arthroplasty*. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S088354031830601X>
3. He, F. J., & MacGregor, G. A. (2011). Dietary Sodium and Potassium Intake and Cardiovascular Risk. *European Heart Journal*. Retrieved from https://academic.oup.com/eurheartj/article/32/24/3087/491667