1. **Provide your team background and organization description (if applicable).**

   I am a Computer Vision and Deep Learning Engineer with seven months of experience in the field. My journey began with a solid foundation in Computer Science Engineering, graduating from the National School of Computer Science in Tunisia. Prior to that, I excelled in the Engineering Preparatory Cycle, which led me to join ENSI three years ago.
   My passion for mathematics and research in the field of Artificial Intelligence has been a driving force throughout my academic and professional endeavors. I find great fulfillment in participating in machine learning competitions, where I continuously strive to expand my knowledge and contribute to innovative solutions.

2. **Explain why you participated in the Alim' Confiance challenge.**

   While I have engaged in multiple machine learning contests previously, what particularly excites me about Trutsii is its focus on promoting ethical and impactful AI applications, especially within the critical domain of medicine. This aligns closely with my belief in leveraging technology for the betterment of society, particularly in areas as crucial as healthcare. Additionally, Trutsii's emphasis on enhancing consumer experiences underscores its commitment to practical and meaningful outcomes, further emphasizing the significance of its competitions.

3. **Describe how you built your winning model and elaborate on the technical and modeling choices you made.**

   In building my winning model, I prioritized addressing key challenges within the dataset, focusing primarily on data imbalance, missing values, and feature engineering. My approach began with meticulous data preprocessing, where I tackled each challenge systematically.
   For missing values, I explored various methods outlined in the submitted notebook, experimenting with different approaches to determine the most effective means of imputation. Additionally, I undertook extensive feature engineering efforts, generating approximately 50 new features to enhance model performance. While not all generated features were utilized in training, this process significantly enriched the dataset.
   To address data imbalance, I experimented with several strategies, including over/undersampling techniques and treating minority classes as outliers. However, the most successful approach involved optimizing an XGBoost model through Optuna, specifically focusing on optimizing the F1 score rather than

accuracy. This adjustment effectively mitigated the imbalance issue, leading to improved model performance.
Furthermore, I explored ensemble methods, experimenting with various algorithms such as LightGBM, Gradient Boosting, Random Forest, AdaBoost, and CatBoost. However, empirical results demonstrated limited success with ensembling compared to optimizing the XGBoost model alone or in conjunction with other models.

Overall, my winning model's success stemmed from a comprehensive approach to data preprocessing, strategic optimization of the XGBoost model.

4. What CPU/RAM resources you used to build your model

CPU: Intel Core i7-10870H (16 cores)
RAM: 15 GB