

## 1. Accuracy (Exactitude)

- Formule :

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

où :

- TP : True Positives (Vrai Positifs)
- TN : True Negatives (Vrais Négatifs)
- FP : False Positives (Faux Positifs)
- FN : False Negatives (Faux Négatifs)
- Interprétation** : L'accuracy mesure la proportion de prédictions correctes parmi toutes les prédictions faites. C'est une bonne métrique lorsque les classes sont équilibrées.

## 2. Precision (Précision)

- Formule :

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- Interprétation** : La précision mesure la proportion de prédictions positives correctes. En d'autres termes, parmi toutes les instances prédites comme positives, quelle fraction était réellement positive. Une précision élevée signifie qu'il y a peu de faux positifs.

## 3. Recall (Rappel ou Sensibilité)

- Formule :

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- Interprétation** : Le rappel mesure la capacité du modèle à identifier toutes les instances positives. En d'autres termes, parmi toutes les instances positives réelles, quelle proportion a été correctement identifiée par le modèle. Un rappel élevé signifie que le modèle détecte la majorité des positifs, mais peut avoir plus de faux positifs.

## 4. F1 Score

- Formule :

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Interprétation** : Le F1 score est la moyenne harmonique de la précision et du rappel. Il est utilisé lorsque vous avez un déséquilibre entre les classes et qu'une simple précision ne donne pas une bonne vue d'ensemble. Le F1 score cherche à équilibrer la précision et le rappel.

## 5. ROC AUC (Area Under the Curve)

- **Formule** : L'aire sous la courbe ROC (Receiver Operating Characteristic) est calculée comme l'intégrale de la courbe de performance du modèle, qui représente la relation entre le taux de vrais positifs (True Positive Rate, TPR) et le taux de faux positifs (False Positive Rate, FPR).
- **Interprétation** : Le ROC AUC mesure la capacité du modèle à distinguer entre les classes. Une valeur de 1.0 signifie un modèle parfait, tandis qu'une valeur de 0.5 indique un modèle sans capacité discriminante (modèle aléatoire).

## 6. Biais (Bias)

- **Formule** :

$$\text{Bias} = \frac{\text{FN} + \text{FP}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

- **Interprétation** : Le biais est une mesure de l'erreur systématique dans les prédictions du modèle. Un biais faible signifie que le modèle est équilibré dans ses prédictions et qu'il ne pénalise pas une classe par rapport à une autre.

## Interprétation des résultats dans le tableau :

- **KNN (Heuristique)** :
  - A une précision et un rappel presque identiques, ce qui montre qu'il est bien équilibré pour identifier les positifs et les négatifs. Cependant, l'accuracy et le F1 score sont un peu plus faibles que ceux de la régression logistique.
  - Le **ROC AUC** indique une bonne capacité de discrimination, mais légèrement inférieure à celle des autres modèles.
- **Régression Logistique (Optimisation)** :
  - Le modèle a la meilleure **accuracy** et **F1 score**, ainsi qu'une précision et un rappel proches, indiquant une très bonne performance globale.
  - Le **ROC AUC** est également le plus élevé, ce qui signifie que le modèle est excellent pour distinguer les classes.
  - Le **biais** est extrêmement faible, indiquant un modèle bien équilibré.
- **Random Forest (Machine Learning)** :
  - Ce modèle montre un excellent **rappel**, indiquant qu'il est particulièrement performant pour détecter les tumeurs malignes (positifs). Cependant, la **précision** est légèrement inférieure à celle de la régression logistique.
  - Le **ROC AUC** est également élevé, et le **biais** est modéré.

En résumé, le modèle de **régression logistique** semble être le plus performant dans l'ensemble pour cette tâche, avec un excellent compromis entre précision, rappel et capacité de discrimination.