

# Rapport sur l'analyse des maladies cardiaques

---

Réalisé par :

- *ZAHA Fatima Zahra*
- *TIRAOUI Doha*
- *GUELBAOUI Hajar*
- *ABOUHADID Asmae*
- *AHMEDLAAZIZ Zakariae*

## Introduction

Les maladies cardiaques constituent l'une des principales causes de mortalité dans le monde. La prédiction précoce de ces maladies à l'aide de modèles de classification peut jouer un rôle crucial dans leur prévention. Ce projet explore l'application de différentes méthodes de data mining et d'apprentissage automatique pour prédire la présence de maladies cardiaques à partir de caractéristiques cliniques et biologiques.

## Description explicative des méthodes choisies

### 1. Recursive Feature Elimination (RFE):

- **Avec SVM** : Le RFE est utilisé avec un SVM linéaire comme estimateur pour éliminer de manière itérative les caractéristiques les moins pertinentes. La sélection finale conserve les 5 caractéristiques les plus influentes. Le modèle est ensuite entraîné et évalué avec ces caractéristiques pour prédire la présence d'une maladie cardiaque.
- **Avec Firefly Algorithm** : Le RFE est combiné avec un algorithme Firefly. Cet algorithme bio-inspiré utilise un processus d'attraction pour

explorer des combinaisons optimales de caractéristiques. Les lucioles représentent des sous-ensembles de caractéristiques, et leur luminosité reflète la précision obtenue par un modèle.

- **Avec Naive Bayes** : Ici, le RFE est appliqué pour réduire les caractéristiques, et un modèle Naive Bayes est utilisé pour mesurer la précision obtenue.

## 2. Exhaustive Feature Selection:

- **Avec SVM** : Cette méthode examine toutes les combinaisons possibles de caractéristiques pour trouver celles qui maximisent la performance d'un modèle SVM.
- **Avec Firefly Algorithm** : L'algorithme Firefly est combiné à la recherche exhaustive pour guider le processus de sélection, réduisant le temps de calcul tout en maintenant une exploration détaillée des combinaisons de caractéristiques.
- **Avec Naive Bayes** : Le modèle Naive Bayes est entraîné et évalué sur toutes les combinaisons possibles de caractéristiques.

## 3. Backward Selection:

- **Avec SVM** : Cette méthode commence par inclure toutes les caractéristiques, puis les élimine progressivement une par une, en évaluant l'impact de chaque suppression.
- **Avec Firefly Algorithm** : La méthode combine une sélection à rebours avec l'algorithme Firefly pour détecter les combinaisons optimales tout en évitant les caractéristiques redondantes.
- **Avec Naive Bayes** : Un processus similaire est appliqué, où les caractéristiques sont supprimées une à une en utilisant un modèle Naive Bayes pour évaluer leur pertinence.

#### 4. Forward Selection:

- **Avec SVM** : Les caractéristiques sont ajoutées une par une, en choisissant à chaque étape celle qui améliore le plus la précision d'un SVM.
- **Avec Firefly Algorithm** : L'algorithme Firefly est intégré pour guider la sélection progressive des caractéristiques, en identifiant dynamiquement celles qui augmentent la précision globale.
- **Avec Naive Bayes** : Une stratégie similaire est appliquée pour les modèles Naive Bayes, avec des ajouts progressifs des caractéristiques en fonction de leur contribution à la précision.

#### Méthode avec le score le plus élevé et explication

Les résultats des méthodes montrent que la combinaison de Recursive Feature Elimination avec l'algorithme Firefly a donné le meilleur score, avec une précision de 92.16%.

- Selected Features (Firefly Algorithm): [1, 2, 4, 5, 6, 7, 9, 10]
- Selected Feature Names (Firefly Algorithm): ['sex', 'chest pain type', 'cholesterol', 'fasting blood sugar', 'resting ecg', 'max heart rate', 'oldpeak', 'ST slope']
- Best Accuracy (Firefly Algorithm): 92.16%

Cette performance peut être attribuée à la capacité de l'algorithme Firefly à explorer efficacement l'espace des solutions et à optimiser la sélection des caractéristiques.

#### Conclusion

Ce projet a exploré différentes approches pour la prédiction des maladies cardiaques. Parmi les méthodes étudiées, la combinaison de RFE et de l'algorithme Firefly s'est avérée la plus performante avec une précision de 92.16%. Cela met en évidence le potentiel des algorithmes bio-inspirés pour la sélection des caractéristiques. Les prochaines étapes pourraient

inclure l'amélioration des hyperparamètres, une meilleure gestion des données manquantes, et l'exploration d'autres modèles comme les réseaux de neurones.