

Problématique

Dans le domaine médical, les patients atteints de maladies cardiaques sont confrontés à une variété de symptômes et de situations, ce qui rend souvent difficile leur classification dans des catégories de diagnostic précises.

- Optimisation des méthodes de diagnostic assistées par ordinateur.
- Intégration efficace d'algorithmes de détection des maladies cardiaques
- Assistance à la classification rapide des patients.
- Développement de recommandations de traitements personnalisés.

Objectifs

L'objectif principal du projet est de développer une application de diagnostic médical automatisé visant à simplifier le processus de diagnostic des maladies cardiaques pour les professionnels de la santé.

- Gérer les données médicales.
- Entraîner l'arbre de décision pour le diagnostic.
- Évaluer ses performances.
- Utiliser l'arbre de décision pour un diagnostic rapide et précis.
- Améliorer les soins aux patients.

Méthodologie suivie

La Figure 1 résume les différentes étapes du système proposé.

1. Prétraitement des données

Nettoyage, transformation et mise en forme des données brutes.

2. Chargement CSV [1] [3]

Ajout d'une fonctionnalité pour importer un fichier CSV d'entraînement.

3. Entraînement modèle

Utilisation des données pour former un modèle d'arbre de décision.

4. Performances du modèle

Présentation des performances du modèle à l'utilisateur, comme le taux de reconnaissance et la matrice de confusion.

5. Interface utilisateur

Ajout d'une fonctionnalité permettant à l'utilisateur de fournir les informations nécessaires pour une prédiction.

6. Prédiction [2]

Réalisation de la prédiction et affichage des résultats à l'utilisateur avec une interface intuitive basée sur Windows Presentation Foundation (WPF).

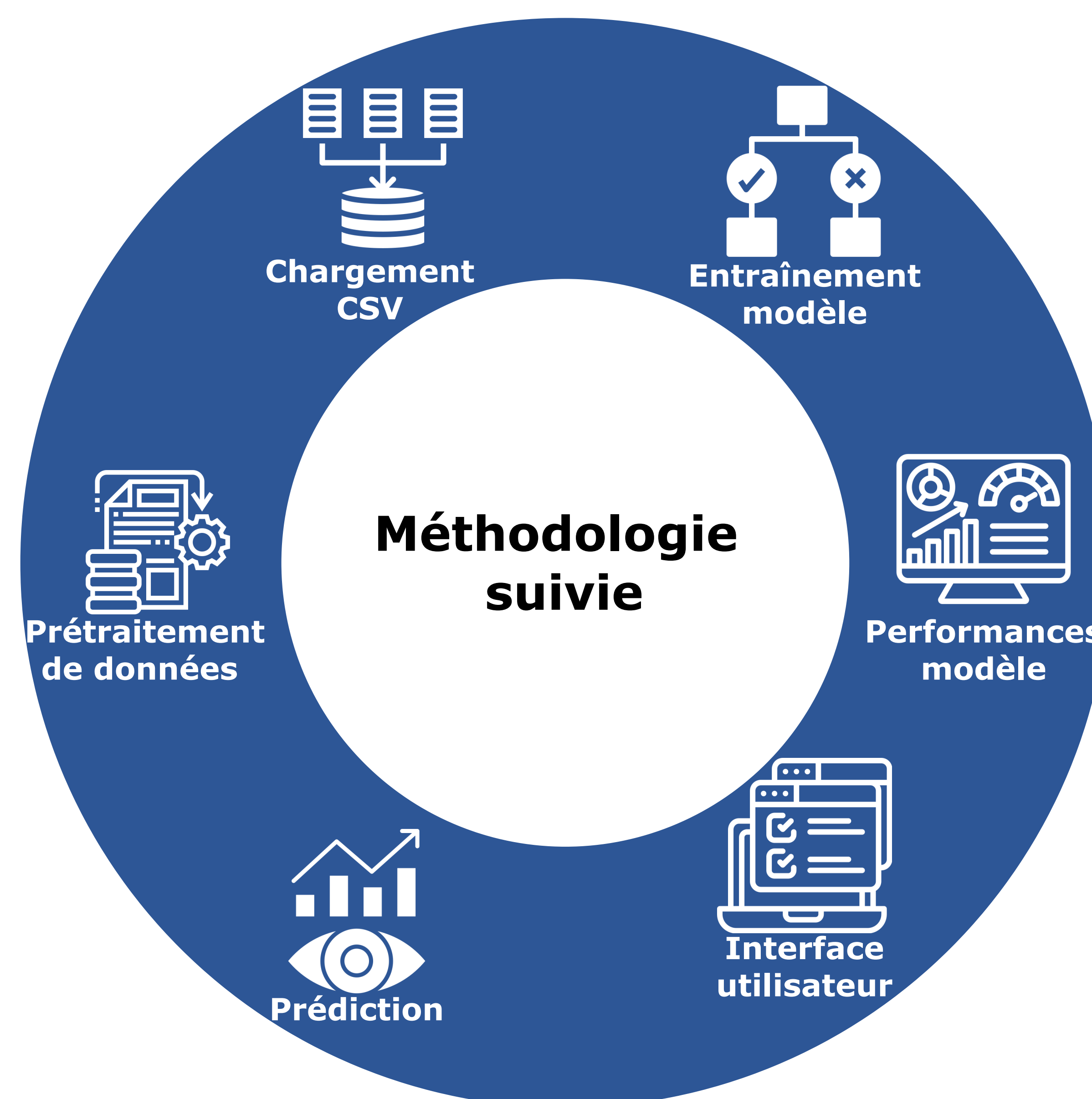


FIGURE 1 – Methodologie.

Évaluation

- Évaluer la performance de l'arbre de décision.
- Adopter une stratégie de validation consistant en la division des données en deux ensembles : ensemble d'apprentissage (67% des données) et de test (33% des données).
- Vérifier la capacité de généralisation du modèle.

Résultats

TABLE 1 – Matrice de confusion sous forme de tableau.[1] [2] [3]

	Prédit Non	Prédit Oui
Vrai Non	161	0
Vrai Oui	3	166

- **Vrais négatifs (VN)** : Cela signifie que 161 exemples de la classe négative (non) ont été correctement classés comme tels.
- **Faux positifs (FP)** : Il n'y a aucun exemple de la classe négative qui a été incorrectement classé comme positif.
- **Faux négatifs (FN)** : Trois exemples positifs ont été mal classés comme négatifs.
- **Vrais positifs (VP)** : 166. Cela signifie qu'il y a 166 exemples de la classe positive qui ont été correctement classés comme tels.
- **Précision** : le taux de reconnaissance global est d'environ 0.989, ce qui correspond à 98.9%.

En résumé, le modèle a très bien fonctionné, avec un petit nombre d'erreurs de classification.

Conclusion

- Ce projet souligne l'importance croissante de la technologie médicale en fournissant des outils avancés pour évaluer et traiter les patients de manière plus efficace et précise.
- En perspective : L'intégration de données en temps réel permettrait une surveillance continue et une détection précoce des anomalies de santé, offrant ainsi de nouvelles possibilités pour améliorer la prise en charge médicale.

Références

- [1] Steinbrunn William Pfisterer Matthias Janosi, Andras and Robert Detrano. Heart Disease. UCI Machine Learning Repository, 1988. DOI : <https://doi.org/10.24432/C52P4X>.
- [2] Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining : Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, Amsterdam, 3 edition, 2011.
- [3] Mangasarian Olvi Street Nick Wolberg, William and W. Street. Breast Cancer Wisconsin (Diagnostic). UCI Machine Learning Repository, 1995. DOI : <https://doi.org/10.24432/C5DW2B>.