



CLERMONT AUVERGNE INP ISIMA

Compte-rendu

La prédiction de maladies cardiovasculaires

Encadrant :

Pr. Bárbara Martín Ruiz

Rédigé par :

Mohammed Oussama LAYAIDI

Le 28 janvier 2024

1 Introduction

L'analyse des données cardiovasculaires est cruciale pour comprendre les facteurs influençant les maladies du cœur. Le code présent vise à explorer un ensemble de données, examinant des caractéristiques objectives et subjectives, ainsi que des paramètres médicaux. Cette investigation permettra d'identifier des tendances significatives pour une meilleure compréhension des relations entre les variables et la prévalence des maladies cardiovasculaires.

2 Analyse et traitement des données

Le processus de préparation des données implique le chargement du jeu de données, suivie d'une conversion et d'une exploration approfondie pour obtenir des informations et des statistiques descriptives essentielles, offrant ainsi un aperçu détaillé des caractéristiques clés et des tendances émergentes, posant ainsi les bases nécessaires pour les analyses ultérieures.

3 Modélisation et évaluation des performances

La section dédiée à la modélisation et à l'évaluation des performances englobe une présentation détaillée des modèles employés, tels que la Régression Logistique, l'Arbre de Décision, le RandomForest, le KNN, le XGBoost et le Réseau de Neurons. Chaque modèle est soumis à un processus d'entraînement, incluant une recherche minutieuse des hyperparamètres optimaux pour maximiser la précision prédictive. L'évaluation des performances est effectuée sur un ensemble de validation distinct, permettant ainsi une comparaison approfondie des résultats obtenus par chaque modèle, afin de sélectionner le plus performant pour la prédiction des maladies cardiovasculaires.

4 Conclusion

En synthèse, l'analyse approfondie des données cardiovasculaires a révélé des tendances significatives, éclairant les relations entre diverses caractéristiques et la prévalence des maladies cardiovasculaires. Le modèle de réseau de neurones, se démarquant par une précision remarquable de 72,7%, a captivé l'attention en tant que choix optimal.

Toutefois, l'atteinte de niveaux de précision encore plus élevés reste un objectif essentiel. Cette aspiration découle de la complexité inhérente à la prédiction des maladies cardiovasculaires, de la nécessité d'intégrer des variables additionnelles pertinentes, et de l'exploration approfondie de techniques avancées d'apprentissage automatique.

En envisageant l'avenir, des améliorations potentielles pourraient inclure l'exploration de nouvelles variables, l'optimisation des hyperparamètres, et l'expansion de l'ensemble de données pour renforcer encore davantage la précision du modèle.