

Analyse

Contexte du Projet

Le projet se concentre sur l'analyse et la classification des données relatives au cancer du sein, une préoccupation majeure de santé publique touchant des millions de personnes dans le monde. Face à cette problématique, il est impératif de développer des approches novatrices pour améliorer la détection précoce et la classification précise des tumeurs.

Dans cette optique, notre projet se positionne comme une initiative technologique exploitant les avancées de l'apprentissage automatique dans le but d'optimiser la précision diagnostic du cancer du sein. Au cœur de notre démarche réside une analyse approfondie d'un jeu de données exhaustif, englobant diverses caractéristiques liées au cancer du sein. Ces caractéristiques incluent des mesures telles que le rayon moyen, la texture, le périmètre, la zone, la régularité, la compacité, la concavité, les points concaves, la symétrie, et d'autres paramètres significatifs. Notre approche d'analyse de données vise à extraire des informations pertinentes de ces caractéristiques, explorant les relations complexes entre elles pour parvenir à une compréhension approfondie des différents types de tumeurs. Dans le contexte de ce projet, l'utilisation de techniques d'apprentissage automatique revêt une importance particulière. Ces techniques permettent de développer des modèles capables de classer efficacement les tumeurs comme étant malignes (M) ou bénignes (B). Cela nécessite une exploration minutieuse des schémas et des tendances présents dans les données, ouvrant la voie à une prise de décision automatisée basée sur des critères spécifiques. Il est également essentiel de souligner que notre projet accorde une attention particulière à l'interprétabilité des résultats obtenus. Les modèles générés doivent être compréhensibles par les professionnels de la santé, favorisant ainsi une adoption plus large et une confiance accrue dans les décisions diagnostiques automatisées.

Problématique

Les méthodes traditionnelles de diagnostic du cancer du sein peuvent présenter des limites en termes de précision et de rapidité. Les professionnels de la santé sont confrontés au défi de traiter un grand volume de données, souvent complexes, pour établir des diagnostics fiables. Dans ce contexte, l'intégration de techniques d'apprentissage automatique offre une opportunité de relever ces défis en exploitant pleinement le potentiel des données disponibles.

Idée du Projet

L'idée principale du projet est de développer un modèle d'apprentissage automatique capable de classer de manière précise les tumeurs du sein comme malignes ou bénignes en se basant sur les caractéristiques fournies. L'analyse implique le prétraitement des données, la visualisation, la gestion des valeurs aberrantes, et la formation de modèles d'apprentissage automatique (en particulier, un classificateur d'arbre de décision et un classificateur Naïve Bayes).

Objectifs

L'objectif principal de ce projet est de développer un modèle d'apprentissage automatique capable de classer avec précision les tumeurs du sein comme étant malignes ou bénignes. Pour atteindre cet objectif, le projet se concentre sur l'exploration approfondie des données, la visualisation des caractéristiques pertinentes, la gestion des valeurs aberrantes, et la formation de modèles de Machine Learning, notamment un classificateur d'arbre de décision et un classificateur Naive Bayes. L'évaluation des performances des modèles se fera à l'aide de mesures telles que la précision, le rappel et le score F1.