SDG 3: Good Health and Well-Being

Report Summary

SDG Addressed:
This project focuses on **SDG 3: Good Health and Well-Being**, specifically addressing the early prediction of **heart disease risk** as a way to promote healthier living and prevent serious health complications.

Problem Statement:
Heart disease is one of the leading causes of death globally. Early identification of individuals at high risk can enable timely lifestyle changes and medical interventions. This project uses machine learning to predict the likelihood of heart disease based on patient health data.

Machine Learning Approach:
We implemented a **supervised learning** approach using classification algorithms. Models such as **logistic regression** and **decision trees** were trained on labeled data to predict whether a person is likely to develop heart disease. The **UCI Heart Disease dataset**, accessed via Kaggle, served as the foundation for training and evaluation.

Dataset and Tools:

- **Dataset**: UCI Heart Disease Dataset (includes features like age, cholesterol levels, blood pressure, etc.)

- **Tools Used**: Python, Jupyter Notebook, Scikit-learn for modeling, and Matplotlib/Seaborn for visualization.

- **Preprocessing**: Data was cleaned, normalized, and split into training and testing sets (typically 80/20 split).

- **Model Evaluation**: Accuracy, precision, recall, and F1-score were used to assess model performance. The logistic regression model achieved approximately **85% accuracy** on the test set.

Ethical Considerations:
Predictive models in healthcare must be handled with caution. **Bias in data**—such as underrepresentation of certain age groups or ethnicities—can result in inaccurate predictions for those populations. Moreover, predictions should support—not replace—clinical decision-making. To ensure **fairness and sustainability**, future improvements could include using more diverse datasets, explaining model outputs clearly to users, and prioritizing patient privacy and consent when handling health data.