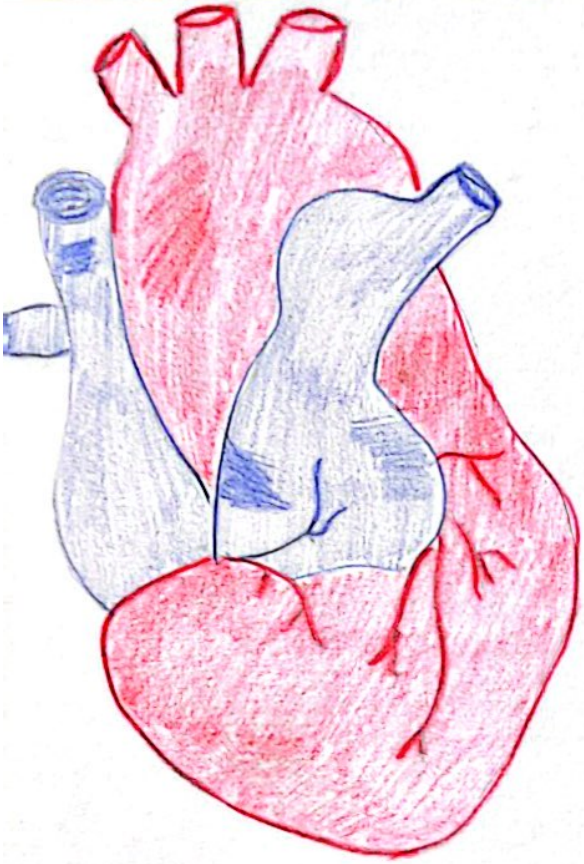
Predicting Health Risks Stay Healthy Stay Ahead!

Dataset: heart_attack_prediction_dataset.csv

Problem Statement

Health issues like heart disease, diabetes and obesity are on a rise and understanding how life choices impact these conditions are more important than ever. This matters as early prediction of health risks is crucial and can help people make better choice and lead healthier lives. It is commonly known in medicine that prevention is better than cure!



Methodology (steps):

- Data Cleaning: Handle missing values and standardize lifestyle-related variables.
- Exploratory Data Analysis (EDA): Explore correlations between lifestyle factors (e.g. smoking, exercise) and health outcomes (focus on heart disease).
- Feature engineering: Create aggregated health indices, lifestyle scores and demographic profiles.
- Modeling: Implement logistic regression and KNN classification to predict health risks. Use tree-based models to predict health risks and understand the impact of various lifestyle factors.
- Evaluation: Use metrics like precision, recall and ROC AUC for model evaluation.
- Insights: Identify key lifestyle factors associated with health risks and suggest preventive measures.

Best Model

XGBoost

- Accuracy: 59.4%
- Precision: 38.9%
- Recall: 23.2%
- f1 score: 29.1%
- Log Loss: 0.768

XGBoost gave us the best performance overall. It handled complex patterns better than other models and gave one of the highest Precision and lower log loss. Lastly it had a good recall meaning it made the positive predictions with confidence.