Heart Attack Classification

Dataset: Heart Attack Analsysis Prediction Dataset

Objective: To develop an accurate predictive model that can help healthcare professionals identify individuals at risk of a heart attack, allowing for early intervention, by predicting and classifying heart attacks using Random Forest Classification on the chosen dataset.

How will the project be doing it?

1. *Data Analysis and Visualization*: The first step will involve conducting a thorough analysis of the dataset. This phase will involve data cleaning, exploration, and feature engineering to understand the relationships and patterns that may exist, in order to prepare the data to apply the appropriate model.
2. *Feature Selection*: After the analysis has been done, the next step will involve selecting the most relevant features provide a two-fold measure - to ensure accuracy of the model and reduce computational complexity.
3. *ML Model*: Once the selection is done, the next step would be to implement the **Random Forest Algorithm**, due to its ability to handle complex classification tasks. This algorithm has been chosen over single decision tree as it is less prone to overfitting complex datasets and improves generalization for untested and unseen data. The model will also be fine-tuned by optimizing hyperparameters.
4. *Validation and Testing*: In the final step, performance will be evaluated using various metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. This will help us ensure the model's ability to predict heart attacks accurately.

What will be the output?

The primary output of this project will be a robust machine learning model capable of accurately classifying individuals as either at risk of a heart attack or not. The Jupyter notebook, with the code, as well as prediction on untested data will also be submitted.