Project Report:

Heart Disease Risk Prediction System

1. Introduction

This exam task, I focused on developing a system to assess heart disease risk. I observed that manually checking patients' risk levels can be slow and prone to error, so I designed a system that automatically classifies patients into five risk levels, from 0 (lowest risk) to 4 (highest risk), using clinical data collected during patient visits. The purpose of this task was to provide medical staff with a more detailed and accurate assessment than just labeling patients as "sick" or "healthy." This system helps improve decision-making, prioritizes high-risk patients, and supports early intervention, making patient care more efficient and effective.

2. My Data Analysis (EDA)

I started with a dataset of **5,000 patients**. Each patient has 13 features like age, blood pressure, and cholesterol.

- I checked for missing values and handled them so the model wouldn't get confused.
- I looked at the "Target" (the risk level) and found that I needed to balance the data so that the "Immediate Danger" cases were just as easy for the model to find as the "No Disease" cases.
- I used charts to see which features matter most—I found that chest pain type and the number of major vessels were huge indicators of high risk.

3. My Technical Process

3.1 Preparing the Data

I created a robust "Pipeline" to process the data automatically:

- **Missing Data:** I filled in gaps using the median value for numbers.
- **Scaling:** I scaled the features (like blood pressure) so they all use the same range. This helps the AI learn faster.
- **Encoding:** I turned text-based categories into numbers.
- **Splitting:** I split my data into 80% for training and 20% for testing. I used **stratification** to make sure the risk levels were spread evenly between both groups.

3.2 Training the Model

I didn't just pick one model; I tested several to find the best one for our hospital:

1. **Random Forest** (This was my winner!)

2. **Gradient Boosting**
3. **SVM and KNN**
4. **ANN (Neural Networks)**

4. Results and Evaluation

My **Random Forest** model was the best fit. I evaluated it using a Confusion Matrix to make sure it wasn't misclassifying high-risk patients as low-risk.

- **Accuracy:** My model achieved high accuracy on the test set.
- **Clinical Safety:** I focused on **Recall** for Class 4 (Immediate Danger) because, in a hospital, we cannot afford to miss a critical patient.

5. Deployment

I didn't just stop at a notebook; I made the system usable for real doctors:

- **Flask API:** I built a backend server that takes patient data and gives a prediction in less than a second.
- **Web Interface:** I created a clean, responsive website using HTML and Bootstrap.
- **Color Coding:** I made the results visual. If a patient is at high risk, the screen turns **Red**. If they are healthy, it stays **Green**. This makes it easy for staff to react quickly.

6. Conclusion

I have successfully built a complete system that goes from raw data to a working hospital tool. It handles data cleaning, smart prediction, and a user-friendly interface. This system is now ready to help the team at CHUB save lives by identifying high-risk heart patients faster.