1.Name

-Rubel Ahmod

2.Student ID (ID that is related to the course)

-2

3.Project Title

- Heart Disease Prediction

4.Provide a brief description of your project (2-3 sentences).

- This project aims to predict the presence of heart disease using patient health data. Various machine learning models are used to analyze medical attributes and determine whether a person is at risk.

5.What is the primary problem your project aims to solve?

- To help in early detection of heart disease based on patient health data, allowing for timely medical intervention.

6.Who are the stakeholders or beneficiaries of your project?

- Doctors, Patients, Healthcare institutions, Medical researchers

7.What dataset(s) did you use? Provide the source(s).

-Dataset: **Heart Disease Dataset**
-Source: Kaggle

8.How did you preprocess and clean the data?

-Handling Missing Values using different imputation techniques like SimpleImputer, KNNImputer, and IterativeImputer.

-Feature Scaling using StandardScaler and MinMaxScaler.

-Label Encoding for categorical data.

9.Did you face any missing data or outliers? How did you handle them?

-Missing Data**:** Handled using multiple imputation techniques.
-Outliers**:** No specific details found in the notebook about handling outliers.10.What exploratory

10.data analysis (EDA) techniques did you apply?

- Heatmaps (correlation analysis), Pair plots, Seaborn & Plotly visualizations

11.Did you create any new features? If yes, explain their significance.

-No

12.Which machine learning models or techniques did you use? Why?

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)
- Decision Tree
- Random Forest
- AdaBoost & Gradient Boosting

13.Describe your model training and validation approach (e.g., train-test split, cross- validation).

- Train-test split: 80% training, 20% testing.
- Cross-validation used for better generalization.

14.What performance metrics did you use to evaluate your model?

-Accuracy ScorePrecision, Recall, F1-Score, Confusion Matrix

15.What were the key challenges in model selection and training?

- Imbalanced Data: Some models may struggle with class imbalance.
- Overfitting in Decision Trees: The model might perform well on training data but poorly on test data.
- Hyperparameter Tuning: Required optimization for better model performance.-In the Decision tree that the training error is 0, while the test error is 0.58 so it is an overfitted model.

-In random forest model it has less error in the test data

16.What were your model's final results (accuracy, precision, recall, F1-score, RMSE, etc.)?

- Exact results are not extracted, but the models were evaluated based on accuracy**,** precision**,** recall**,** and F1-score.

17.Did you deploy your model (e.g., Flask API, Streamlit app)? If yes, provide details.

-Yes

18.Provide a link to your GitHub repository Interpretation and Insights

- https://github.com/rubelahmod/Heart_desease_detection.git

19.What key insights did you gain from your project?

- Learned how different health features relate to heart disease.

- Understood how different models perform on this type of data

20.What are the strengths and limitations of your approach?

- Strength: Used multiple models to compare results.

-Limitation: Might need better handling of class imbalance

21.If you had more time, what would you improve or add to this project?

- Try more models like XGBoost.

-Improve dataset size for better predictions.