Assessment Report on: Predict Herat Disease

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

CSE – AI&ML

Suman Verma

Roll no.202401100400194

Under the supervision of

Abhishek Shukla

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

May, 2025

Heart Disease Prediction Report Instructions

Title Page

- **Project Title: *"Heart Disease Prediction Using Machine Learning"***
- **Name: Suman Verma**
- **Date: 18/04/2025**
- **KIET group of institution**

Abstract

- The goal ("*predicting heart disease"*)
- The dataset used
- The methods (e.g., Random Forest)
- Key results (accuracy, F1-score)
- Significance (e.g., impact on early diagnosis)

INTRODUCTION

- Early detection of heart disease through predictive analysis can significantly improve treatment outcomes and prevent complications.
- Machine learning models help forecast patient risks based on medical history, lifestyle, and genetic factors, enabling *preventive care.*
- The *primary objective* of this project is to develop a machine learning model that can accurately predict the presence of heart disease in patients based on various medical attributes. By analyzing patterns in health-related data such as age, blood pressure, cholesterol levels, and ECG results, the model aims to assist in early diagnosis and improve clinical decision-making.

Dataset Description

- Source we use, UCI Heart Disease dataset, Electronic Health Records (EHR), Wearable Devices & Sensors, Machine Learning Models (*AI-driven analysis uses structured datasets to predict risks*)
- Number of samples and features, sample size datasets range from *hundreds to thousands* of patient records.
- Target variable (0 = No disease, 1 = Disease)
- List key features (age, cholesterol, max heart rate, etc.)

Data Preprocessing

- Missing value handling - Missing value handling ensures data integrity by using techniques like imputation, deletion, or interpolation to maintain model accuracy and reliability
- Encoding categorical variables (e.g., chest pain type)
- Feature scaling (e.g., Standard Scaler)
- Train-test split (e.g., 80-20 split)

Results

- Metric scores (put them in a table)

Metric	Description
Accuracy	Overall correctness of predictions.
Precision	Correct positive predictions out of all predicted positives.
Recall (Sensitivity)	Ability to identify actual positives correctly.
F1 Score	Harmonic mean of precision and recall.
ROC-AUC	Measures model discrimination between classes.

Conclusion

In this project, we successfully built a machine learning model to predict the likelihood of heart disease based on a set of medical attributes. Through data preprocessing, feature analysis, and model evaluation, we demonstrated that machine learning—particularly models like Random Forest and Logistic Regression—can achieve high accuracy and reliability in detecting heart disease. The results highlight the potential of data-driven approaches to support early diagnosis and assist medical professionals in making informed decisions.

First 5 rows of the dataset:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	0	145	233	1	2	150	0	2.3	2	0	2	0
1	67	1	3	160	286	0	2	108	1	1.5	1	3	1	1
2	67	1	3	120	229	0	2	129	1	2.6	1	2	3	1
3	37	1	2	130	250	0	0	187	0	3.5	2	0	1	0
4	41	0	1	130	204	0	2	172	0	1.4	0	0	1	0

Dataset Info:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       303 non-null    int64
 1   sex       303 non-null    int64
 2   cp        303 non-null    int64
 3   trestbps  303 non-null    int64
 4   chol      303 non-null    int64
 5   fbs       303 non-null    int64
 6   restecg   303 non-null    int64
 7   thalach   303 non-null    int64
 8   exang     303 non-null    int64
 9   oldpeak   303 non-null    float64
 10  slope     303 non-null    int64
```