

**Project Report:**

**Heart Disease Prediction Using Machine Learning**

**Submitted to:**

Mr Usama Khan

**Submitted by:**

M Shaharyar Shabbir Bhatti

**ID:**

38465

**Subject:**

Machine Learning

**Department:**

BS-AI(6th Semester)

# Heart Disease Prediction Using Machine Learning

## Abstract

This project explores the application of machine learning algorithms to predict heart disease. Utilizing a dataset of 303 patient records, we compare the performance of logistic regression and K-nearest neighbors (KNN) algorithms in predicting the presence of heart disease.

## Introduction

Heart disease is a leading cause of death globally. Early detection is crucial for effective treatment. This project aims to develop a predictive model that uses medical parameters to assess the likelihood of heart disease.

## Dataset Description

The dataset comprises 303 patient records, each with 14 medical features:

* Age
* Sex
* Chest pain type
* Resting blood pressure
* Serum cholesterol
* Fasting blood sugar
* Resting electrocardiogram results
* Maximum heart rate achieved
* Exercise-induced angina (old peak)
* Slope of the peak exercise ST segment
* Number of major vessels colored by fluoroscopy (ca)
* Thalassemia (thal)

The target variable indicates the presence (1) or absence (0) of heart disease.

## Methodology

### Data Preprocessing

* **Missing Values**: Confirmed absence of missing data.
* **Normalization**: Deemed unnecessary as data is well-formatted.
* **Feature Selection**: All features retained due to their relevance.

### Data Splitting

The data is partitioned into training (80%) and testing (20%) sets.

### Model Training

Two models are trained:

* Logistic Regression
* K-Nearest Neighbors (KNN)

### Model Evaluation

Model performance is evaluated based on accuracy scores for both training and testing datasets.

## Results

**Logistic Regression**

* **Training Accuracy**: 85.12%
* **Testing Accuracy**: 81.97%

**KNN**

* **Training Accuracy**: 78.10%
* **Testing Accuracy**: 62.30%

## Discussion

The logistic regression model outperformed the KNN model. The higher accuracy in both training and testing phases suggests logistic regression is more suitable for this dataset.

## Conclusion

Logistic regression is recommended for predicting heart disease in this context due to its superior accuracy and performance over the KNN model.