# **Project Report**

**Heart Disease Prediction** 

December 2023

#### **ISSUED BY**

Dr. Uzair Iqbal

## **REPRESENTATIVE**

**IBRAHIM ABID (21I-0298)** 

# **Table of Contents**

	3
Introduction and Domain Background	3
Introduction	3
Domain Background	3
Dataset Description:	4
Dataset Overview:	4
Attribute Details:	4
Descriptive Statistics:	5
My Contribution:	5

# **Introduction and Domain Background**

### Introduction

Heart disease remains a leading cause of mortality worldwide, emphasizing the critical need for accurate prediction and early intervention. In this report, we are diving into the application of machine learning for heart disease prediction, aiming to contribute to the advancements in cardiovascular healthcare.

# Domain Background

Cardiovascular diseases encompass a range of conditions affecting the heart and blood vessels, including coronary artery disease, heart failure, and arrhythmias. Early detection of heart disease risk factors is paramount for preventive measures and personalized patient care. Machine learning offers a promising avenue for developing predictive models that can assist healthcare professionals in identifying individuals at high risk.

# **Dataset Description:**

#### **Dataset Overview:**

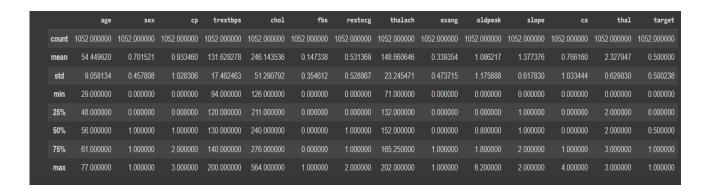
My analysis is based on a comprehensive dataset containing various attributes related to heart health. The dataset comprises numerical features, including age, sex, blood pressure, cholesterol levels, and exercise-induced angina, among others. These attributes provide a rich set of information for training and evaluating machine learning models.

#### Attribute Details:

- Age: The age of the individual.
- Sex: Gender of the individual (0 for female, 1 for male).
- **CP:** Chest pain type (0 to 3).
- Trestbps: Resting blood pressure.
- Chol: Serum cholesterol.
- *Fbs*: Fasting blood sugar > 120 mg/dl (1 for true, 0 for false).
- **Restecg:** Resting electrocardiographic results (0 to 2).
- Thalach: Maximum heart rate achieved.
- **Exang:** Exercise-induced angina (1 for yes, 0 for no).
- Oldpeak: ST depression induced by exercise.
- Slope: Slope of the peak exercise ST segment (0 to 2).
- Ca: Number of major vessels colored by fluoroscopy (0 to 3).
- **Thal:** Thalassemia type (0 to 3).
- *Target*: Presence of heart disease (1 for yes, 0 for no).

# **Descriptive Statistics:**

The table below provides descriptive statistics for the dataset:



# My Contribution:

In this project there were certain flaws which I fixed such as:

- Data Imbalance
- Applied Standard Scaler.
- Printed Classification Report.
- Removed Outliers from dataset.
- Checked for any NaN Values in Dataset.
- Used Random OverSampler to fix imbalanced dataset.
- Used Neural Network with Sigmoid Activation Function at its last Layer.

# **Used Standard libraries:**

# > NumPy:

np.asarray():

Converts input to an array.

np.reshape():

Gives a new shape to an array without changing its data.

## > Pandas:

pd.read\_csv():

Reads a comma-separated values (csv) file into a DataFrame.

DataFrame.head():

Returns the first n rows of a DataFrame.

DataFrame.isna():

Detects missing values.

DataFrame.info():

Prints a concise summary of a DataFrame.

DataFrame.isnull().sum():

Counts the number of missing values in each column.

DataFrame.select\_dtypes():

Selects columns based on their data types.

# > Seaborn:

sns.countplot():

Shows the counts of observations in each categorical bin using bars.

sns.boxplot():

Draws a box plot to show distributions with respect to categories.

# ➤ Matplotlib:

plt.figure():

Creates a new figure.

### plt.show():

Displays the figure.

#### Imbalanced-learn

RandomOverSampler():

Randomly over-samples the minority class.

### Scikit-learn:

### train\_test\_split():

Splits arrays or matrices into random train and test subsets.

## accuracy\_score():

Accuracy classification score.

### StandardScaler():

Standardizes features by removing the mean and scaling to unit variance.

### > TensorFlow-Keras:

## keras.Sequential():

Linear stack of layers for building the neural network model.

# model.compile():

Configures the model for training.

# model.fit():

Trains the model for a fixed number of epochs.

# model.predict():

Generates predictions for input samples.

# Other Python built-in libraries:

os (for file path operations)