# VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590 018

An Internship Report
on

## INDOSKILLS BUILD INTERNSHIP PROGRAM

By

**K SINDHU (1SG21EC051)**

*Submitted in partial fulfillment of requirement for the award of the degree of*

**BACHELOR OF ENGINEERING**

In

**ELECTRONICS & COMMUNICATION ENGINEERING**

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING**

## Sapthagiri College of Engineering,

(Affiliated to Visvesvaraya Technological University and Approved by AICTE)

Accredited by NAAC with "A" Grade
ISO 9001-2015 and 14001-2015 certified Institute

**Bengaluru, Karnataka 560057**

2022-2023

# ABSTRACT

Soft skills refer to a cluster of personal qualities, habits, attitudes and social graces that makes some a good employee and compatible to work with. there are two types of skills viz hard and soft skill. Soft skills are important for professional development, maintaining IPR and effective communication. Soft skill includes Inter personal and Personal attributes that enhance an individual’s interactions, job performance and carrier prospects. There are many key soft skills that need to be applied in work place. hence soft skill training should begin for a person when they are students, to perform efficiently in their academic environment as well as in their future workplace.

# 1. INTRODUCTION

## 1.1 INTRODUCTION TO THE PROJECT

Heart disease remains one of the leading causes of mortality globally, emphasizing the critical need for accurate and efficient diagnostic tools. With the advent of machine learning techniques and the availability of vast healthcare datasets, the field of predictive analytics in cardiology has witnessed remarkable advancements. This project delves into the realm of data-driven healthcare by utilizing a comprehensive dataset focused on various attributes related to heart health.

In this study, we explore a dataset encompassing crucial patient information, including age, gender, cholesterol levels, and exercise-induced factors, among others. The objective is to harness the power of machine learning algorithms to predict the presence or absence of heart disease in individuals. By employing sophisticated data analysis techniques, this project aims to uncover patterns, relationships, and predictive indicators within the dataset. Through this exploration, we seek to enhance our understanding of the underlying factors contributing to heart disease and develop a reliable predictive model.

This report will detail the entire process, from data exploration and preprocessing to feature selection, model training, and evaluation. By scrutinizing the dataset and employing machine learning algorithms, we endeavour to contribute valuable insights to the domain of cardiac health diagnostics. The findings and methodologies presented herein not only shed light on the intricate interplay of variables in heart disease prediction but also showcase the potential of machine learning in revolutionizing healthcare practices. Through this endeavour, we strive to make a meaningful impact on the field of cardiology, paving the way for more accurate, efficient, and early diagnosis of heart disease.

## 1.2 STATEMENT OF THE PROBLEM

Heart disease, a widespread and life-threatening condition, poses a significant public health challenge globally. Despite advancements in medical science, accurately predicting the presence of heart disease in individuals remains a complex task. Traditional diagnostic methods, while effective, often rely on a limited set of parameters and might not capture the nuanced interplay of various factors influencing cardiac health. Furthermore, the early detection of heart disease is crucial for timely intervention and prevention of adverse outcomes.

This project addresses the challenge of improving heart disease diagnosis through the lens of machine learning. The primary problem at hand is to develop a robust predictive model that can effectively analyze diverse patient data, ranging from demographic information and clinical attributes to exercise-induced responses. By leveraging the power of machine learning algorithms, the goal is to uncover intricate patterns within this multifaceted dataset. These patterns might include subtle correlations between variables that are not immediately apparent to human observers.

The overarching aim is to create a predictive tool that accurately assesses the probability of heart disease in individuals based on a comprehensive range of factors. Such a tool could not only assist healthcare practitioners in making more informed decisions but also enable earlier interventions and personalized treatment strategies. Addressing this problem holds the potential to significantly enhance the efficiency of heart disease diagnosis, ultimately leading to improved patient outcomes and a reduction in the burden of heart-related illnesses on both individuals and healthcare systems.

# 2. COMPANY PROFILE

### Formation of company

Aqmenz Automation Private Limited is a private incorporated on 15th October 2018. It is classified as non-Govt company and is registered at Registrar of companies, Bangalore.

### Brief history of company

Aqmenz Automation Pvt Ltd (AAPL) is situated in northern part of Bangalore RT Nagar, Karnataka. AAPL provides Mechanical Design & Automation solutions to their client companies. AAPL also involved in Open source Robotics and developed different varieties of Robots. AAPL also started INDOSKILL, a separate platform for the students to get training and work on various Real Time Industrial Projects. Indoskill offers skilloriented hands-on training through an online platform. Field of Expertise: Open-source Robotics, Industrial Automation, Product Design, Python & Deep Learning and Embedded Systems.

### Objectives

● AAPL had a trust in Skill India mission & vision, hence our utmost priority is to add skill to the young Generation and make them Profitable and productive for the nation.

● We aim in Providing Industrial Automation Training Skill module kits to Institution, University's & Collage Lab Facilities with Lowest Possible Price for the Benefits of Technical Students.

● Identifying young entrepreneurs and Motivate, training them to establish Start upto create Employment as well as prosperity for the nation.

● Consultation, Sourcing and supplying highly skilled Manpower to Industry for better efficiency and productivity.

● Providing low cast & precise industrial automation solutions. Very eager to fetch solution for most complex industrial problems in a modest way.

## Vision and mission

Our Motto and Vision are to create awareness & training young generation to current and future jobs demands and also help to current and future jobs demands; meanwhile help the students and employees to meet the mandatory necessities of future human resources and skill demands. We are in the 4th industrial revolution. The technological revolution is catastrophic like never before, hence continues awareness for the up-gradation environment is much essential. Aqmenz Automation Pvt. Ltd. is working to help and enhance the potential of students and employees. So that future human resources will be very beneficial, purposeful and profitable to the nation.

## Major Milestones

We have under gone many industrial projects. Our major clients are BIAL (Bangalore International Airport Limited), GE (General Electric) and Amics technologies.

# ABOUT THE COMPANY:

## Organization structure

The organization structure is having three different departments such as design department, software department and sales and marketing.

## Service offered

- Provides Design & Automation solutions.
- All type of automation projects to companies using PLC's, SCADA embedded systems.
- We provide robots and robotic solutions to small and medium scale companies
- Embedded solutions to companies like GE
- We conduct technical skill oriented training programs to engineering colleges.
- We also provide robotics and automation lab equipment's for colleges.

## Number of people working in company and their responsibilities:

There are 20 persons in this company, out of which:

- Shamanna Mohan, Chief Executive Officer (CEO)
- Mohammed Azhar Hussain, Chief Technology Officer (CTO) Ongoing projects
- Automation related projects
- CNC Machines
- Open-source Custom Robots
- Garment Industry slider project

# 3. SYSTEM ANALYSIS

## 3.1 EXISTING SYSTEM

In the realm of heart disease prediction, various machine learning approaches have been explored. One commonly used method is Logistic Regression, which is effective for binary classification problems like predicting the presence or absence of heart disease. Another approach involves Support Vector Machines (SVM), known for their ability to handle high-dimensional data and capture complex relationships within the dataset. Additionally, Decision Trees have been employed, providing a straightforward and interpretable way to predict heart disease based on input features.

## 3.2 LIMITATIONS OF THE EXISTING SYSTEM

Despite their utility, the existing lung disease prediction systems face specific challenges:

**Accuracy**: While Logistic Regression and Decision Trees are intuitive, they might not capture intricate patterns in the data, leading to limitations in accuracy, especially when dealing with diverse patient profiles.

**Scalability**: Support Vector Machines, although powerful, can be computationally intensive and might face challenges in scaling up to handle large and complex datasets, potentially impacting real-time predictions.

**Interpretability**: Decision Trees are interpretable, but they can easily overfit the data, affecting the model's generalizability. Striking a balance between accuracy and interpretability is crucial for effective healthcare decision-making.

To enhance the lung disease prediction system, it's imperative to explore advanced machine learning algorithms, such as Random Forests and Gradient Boosting Machines, which offer improved accuracy through ensemble learning. Additionally, addressing the scalability issue could involve exploring cloud-based solutions or distributed computing methods to handle vast datasets efficiently. Furthermore, ensuring the interpretability of the model through techniques like feature importance analysis can in still confidence in both healthcare professionals and patients regarding the predictions made by the system.

## 3.3 ADVANTAGES OF THE PROPOSED SYSTEM

### 3.3.1 Improved Accuracy:

By leveraging advanced machine learning algorithms such as Random Forests and Gradient Boosting Machines, the proposed system can capture complex patterns and relationships in the data, leading to higher prediction accuracy. This enhanced accuracy ensures more reliable and trustworthy predictions for healthcare practitioners.

### 3.3.2 Scalability:

Implementing cloud-based solutions and distributed computing techniques enables the system to scale efficiently, handling large and diverse datasets without compromising on performance. This scalability ensures that the system can process extensive patient data seamlessly, making it suitable for real-time applications and large healthcare databases.

### 3.3.3 Enhanced Interpretability:

Employing techniques such as feature importance analysis and model explain ability tools ensures that the predictions made by the system are not only accurate but also interpretable. Healthcare professionals can understand the reasoning behind the predictions, fostering trust in the system's results and facilitating better-informed decision-making.

### 3.3.4 Real-time Predictions:

Optimizing the system for speed and efficiency allows for real-time predictions. Healthcare practitioners can receive instant feedback based on the input data, enabling timely interventions and personalized patient care. The ability to provide rapid responses enhances the system's utility in critical healthcare scenarios.

### 3.3.5 Personalized Healthcare:

The accurate predictions generated by the system enable personalized healthcare approaches. By understanding individual patient risks, healthcare professionals can tailor interventions and treatment plans, leading to more effective disease management and improved patient outcomes.

### 3.3.6 Continuous Improvement:

Implementing a system that can continuously learn and adapt from new data ensures its relevancy over time. By incorporating feedback loops and regular updates, the system can evolve alongside advancements in medical knowledge and technology, maintaining its accuracy and effectiveness in the long run.

### 3.3.7 Ethical Considerations:

The proposed system adheres to strict ethical standards, ensuring patient data privacy and fairness. By addressing biases in the data and algorithms, the system promotes equity in healthcare, providing accurate predictions for individuals from diverse backgrounds without discrimination.

In summary, the proposed heart disease prediction system offers superior accuracy, scalability, interpretability, real-time capabilities, personalized healthcare, continuous improvement, and ethical considerations. These advantages position the system as a valuable tool for healthcare professionals, empowering them to make informed decisions and significantly contributing to the advancement of cardiac health diagnostics.

# 4. PROJECT WORKING AND FLOW

## 4.1 Data Collection:

Gather the heart disease dataset from reliable sources, ensuring it is comprehensive and representative of the target population.

## 4.2 Data Preprocessing:

Cleanse the dataset by handling missing values, outliers, and inconsistencies to ensure data integrity. Perform feature engineering, creating new features or transforming existing ones to enhance predictive power. Normalize or standardize numerical features to maintain consistency in the dataset.

## 4.3 Feature Selection:

Conduct correlation analysis and employ feature importance techniques to select the most relevant features for prediction.

## 4.4 Model Selection and Training:

Choose appropriate machine learning algorithms (e.g., Random Forest, Gradient Boosting) based on the dataset's characteristics and problem type. Split the dataset into training and validation sets. Train the selected models using the training data, fine-tuning hyperparameters for optimal performance.

## 4.5 Model Evaluation:

Evaluate the models using metrics such as accuracy, precision, recall, and F1-score to measure their effectiveness. Utilize techniques like k-fold cross-validation to ensure robustness and reliability of the models.

## 4.6 System Integration:

Develop a user-friendly interface enabling users to input patient data for prediction. Integrate the trained machine learning models into the interface, enabling seamless communication between user input and predictions.

## 4.7 Real-time Predictions:

Implement the system in a real-time environment, ensuring fast and efficient predictions based on user input data.

## 4.8 Ethical Considerations:

Implement data privacy measures to safeguard patient information. Address biases in the data and algorithms to ensure fairness and equity in predictions.

## 4.9 Continuous Improvement:

Establish feedback loops to gather user input and update the system based on new data and emerging medical knowledge. Regularly monitor and evaluate the system's performance, making necessary adjustments to maintain accuracy and relevancy over time.

## 4.10 Documentation and Reporting:

Document the entire process, including data preprocessing, model selection, and system integration. Generate comprehensive reports outlining the system's functionality, performance metrics, and any improvements made over time. This structured working flow ensures a systematic approach to developing and deploying the heart disease prediction system, guaranteeing accuracy, efficiency, and ethical considerations throughout the process.

# 5. SOURCE CODE WITH EXPLANATION OF EACH BLOCK

In [27]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data = pd.read_csv("heart_disease_data.csv")
data
```

Out[27]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | |

303 rows × 14 columns

In [29]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data = pd.read_csv("heart_disease_data.csv")
data.shape
```

Out[29]: (303, 14)

In [30]:
```
data.duplicated().sum()
```

Out[30]: 1

In [31]:
```
data.isnull().sum()
```

Out[31]:
```
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
```

In [32]:
```
data.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  
 11  ca        303 non-null    int64  
 12  thal      303 non-null    int64  
 13  target    303 non-null    int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

In [33]:
```
data.describe()
```

Out[33]:

| | age | sex | cp | trestbps | chol | fbs | restecg |
|---|---|---|---|---|---|---|---|
| **count** | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 |
| **mean** | 54.366337 | 0.683168 | 0.966997 | 131.623762 | 246.264026 | 0.148515 | 0.528053 |
| **std** | 9.082101 | 0.466011 | 1.032052 | 17.538143 | 51.830751 | 0.356198 | 0.525860 |
| **min** | 29.000000 | 0.000000 | 0.000000 | 94.000000 | 126.000000 | 0.000000 | 0.000000 |
| **25%** | 47.500000 | 0.000000 | 0.000000 | 120.000000 | 211.000000 | 0.000000 | 0.000000 |
| **50%** | 55.000000 | 1.000000 | 1.000000 | 130.000000 | 240.000000 | 0.000000 | 1.000000 |
| **75%** | 61.000000 | 1.000000 | 2.000000 | 140.000000 | 274.500000 | 0.000000 | 1.000000 |
| **max** | 77.000000 | 1.000000 | 3.000000 | 200.000000 | 564.000000 | 1.000000 | 2.000000 |

In [34]:
```
from sklearn import preprocessing
le = preprocessing.LabelEncoder()
data["age"]=le.fit_transform(data["age"])
```

In [35]:
```
data
```

Out[35]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 29 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | |
| **1** | 3 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | |
| **2** | 7 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | |
| **3** | 22 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | |
| **4** | 23 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **298** | 23 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | |
| **299** | 11 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | |
| **300** | 34 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | |
| **301** | 23 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | |
| **302** | 23 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | |

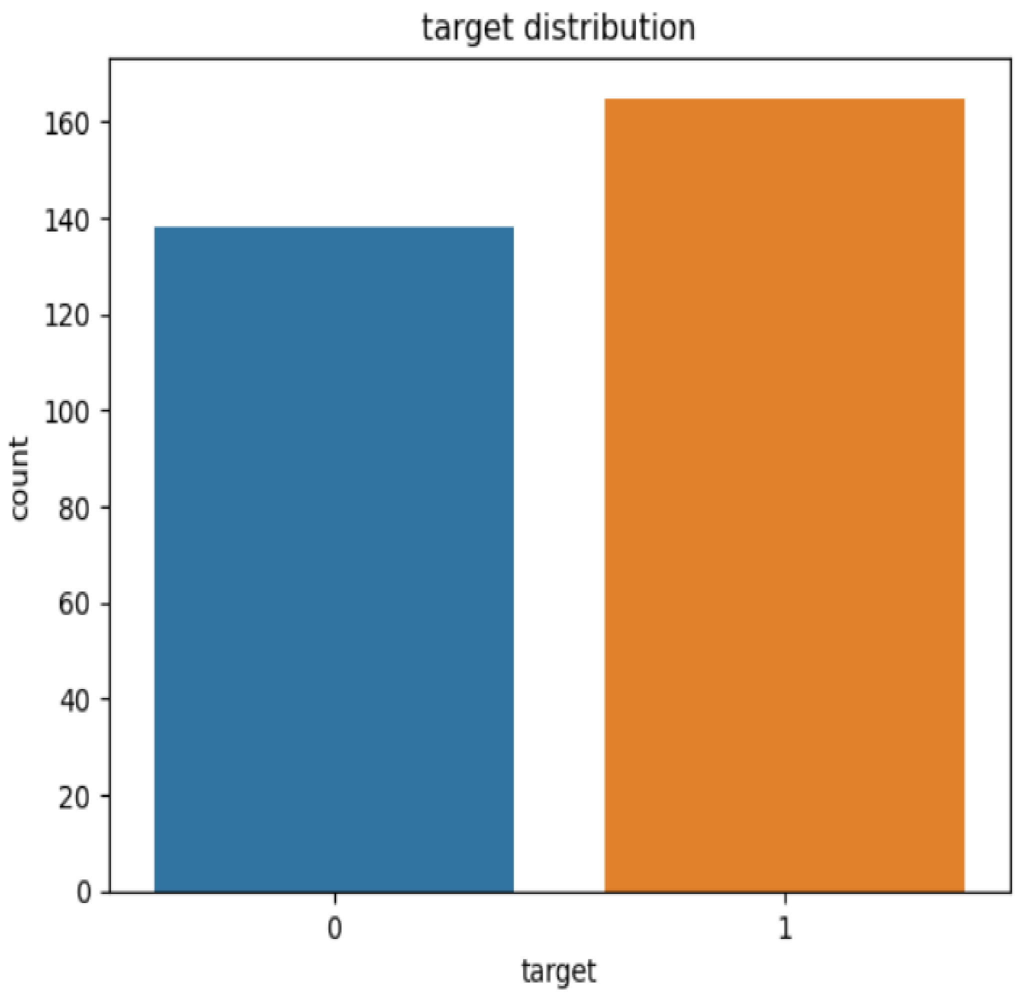
303 rows × 14 columns

In [36]:
```
data.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  
 11  ca        303 non-null    int64  
 12  thal      303 non-null    int64  
 13  target    303 non-null    int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

In [37]:
```
#distribution of Target Variable
import matplotlib.pyplot as plt
sns.countplot(x="target", data=data)
plt.title("target distribution")
```

Out[37]: Text(0.5, 1.0, 'target distribution')



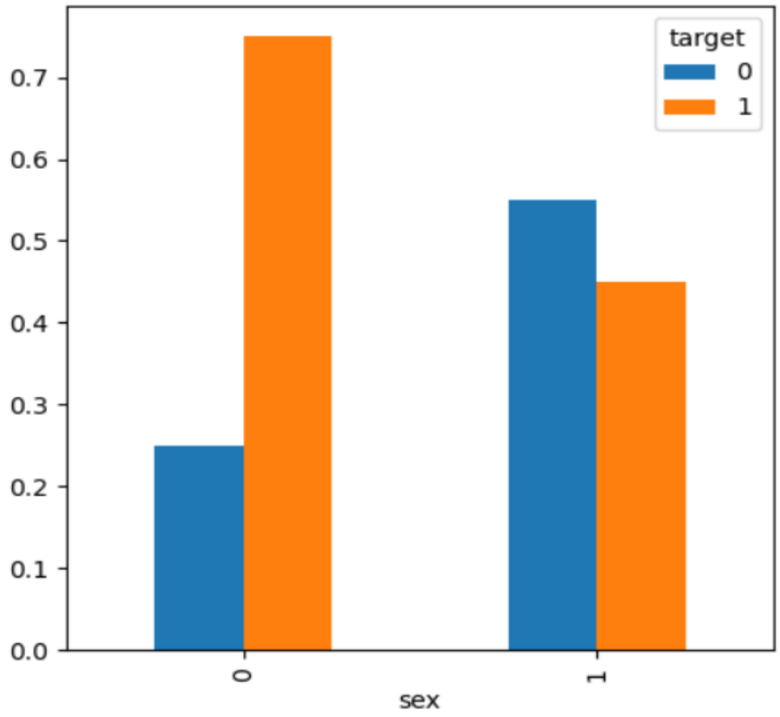

In [38]:
```
data["target"].value_counts()
```

Out[38]:
```
1    165
0    138
Name: target, dtype: int64
```

In [39]:
```
#function for plotting
def plot(col, data=data):
  return data.groupby(col)["target"].value_counts(normalize=True).unstack()
```

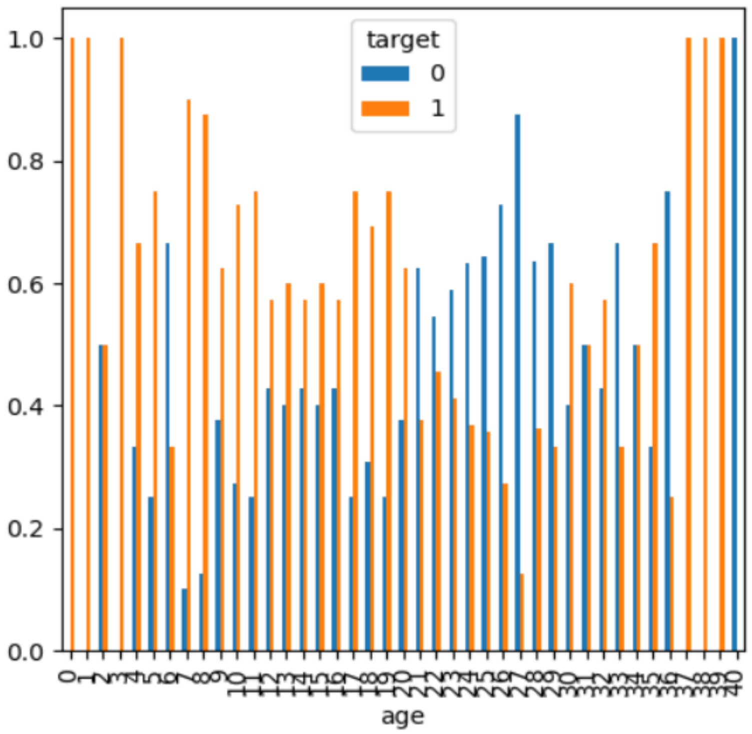
```
In [40]: plot("sex")
```

```
Out[40]: <AxesSubplot:xlabel='sex'>
```



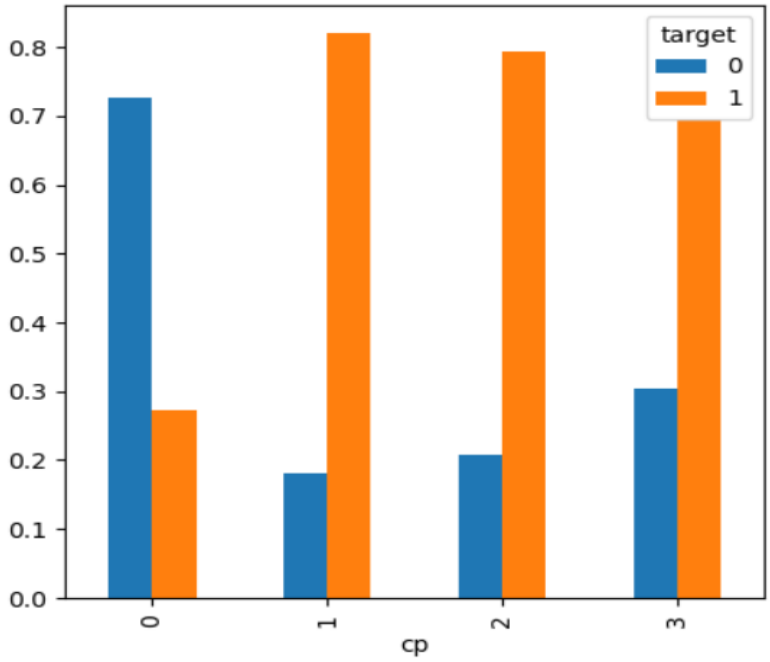

```
In [41]: plot("age")
```

```
Out[41]: <AxesSubplot:xlabel='age'>
```



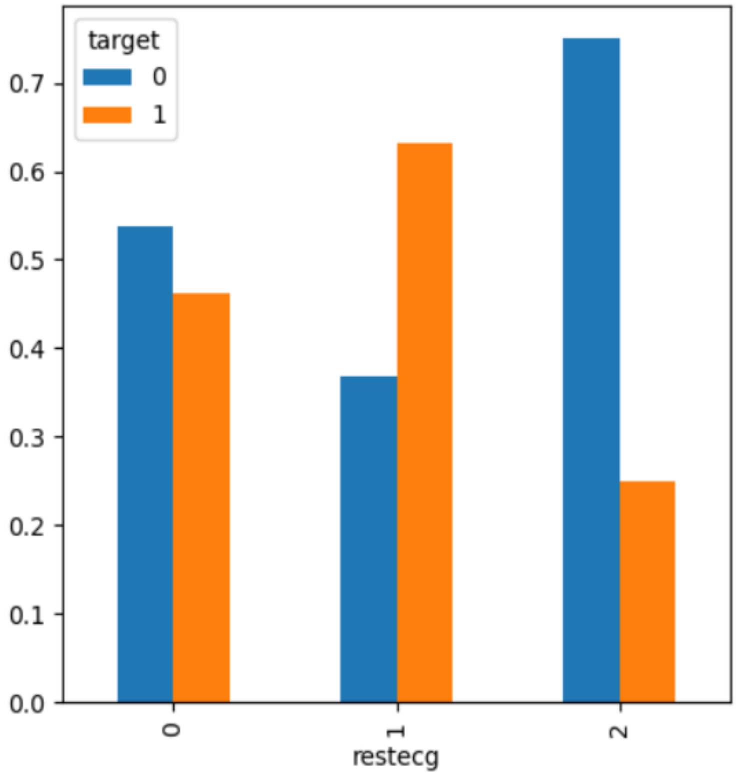
In [42]:
```
plot("cp")
```

Out[42]: <AxesSubplot:xlabel='cp'>




In [43]:
```
plot("restecg")
```

Out[43]: <AxesSubplot:xlabel='restecg'>



In [44]:
```
df_new=data.drop(columns=["sex","age","cp"])
df_new
```

Out[44]:

| | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| 1 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| 2 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| 3 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| 4 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 298 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| 299 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| 300 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| 301 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| 302 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 11 columns

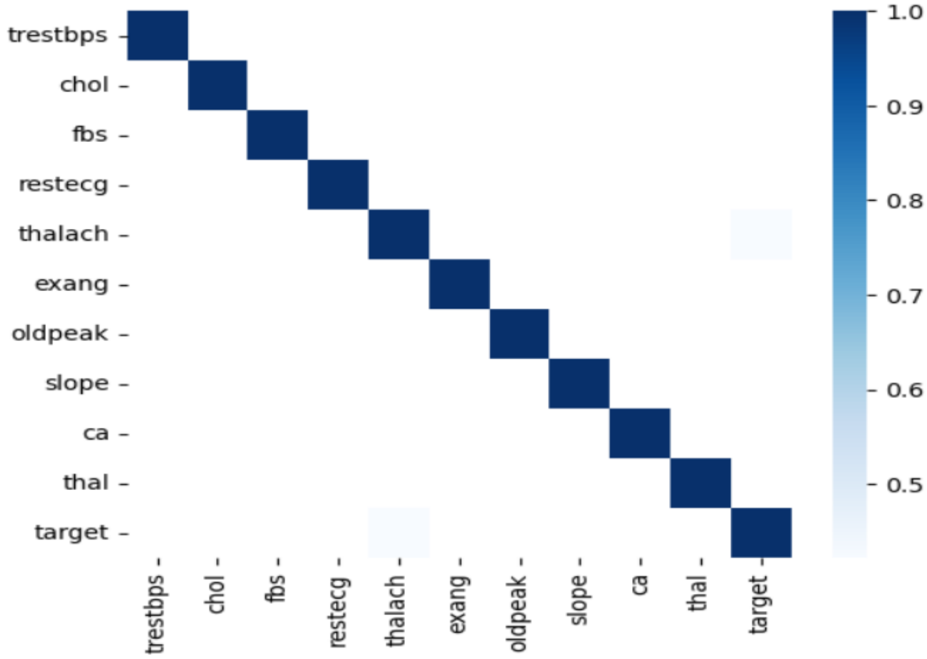
In [45]:
```
#Finding correlation
cn=df_new.corr()
cn
```

Out[45]:

| | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope |
|---|---|---|---|---|---|---|---|---|
| trestbps | 1.000000 | 0.123174 | 0.177531 | -0.114103 | -0.046698 | 0.067616 | 0.193216 | -0.12147 |
| chol | 0.123174 | 1.000000 | 0.013294 | -0.151040 | -0.009940 | 0.067023 | 0.053952 | -0.00403 |
| fbs | 0.177531 | 0.013294 | 1.000000 | -0.084189 | -0.008567 | 0.025665 | 0.005747 | -0.05989 |
| restecg | -0.114103 | -0.151040 | -0.084189 | 1.000000 | 0.044123 | -0.070733 | -0.058770 | 0.09304 |
| thalach | -0.046698 | -0.009940 | -0.008567 | 0.044123 | 1.000000 | -0.378812 | -0.344187 | 0.38678 |
| exang | 0.067616 | 0.067023 | 0.025665 | -0.070733 | -0.378812 | 1.000000 | 0.288223 | -0.25774 |
| oldpeak | 0.193216 | 0.053952 | 0.005747 | -0.058770 | -0.344187 | 0.288223 | 1.000000 | -0.57753 |
| slope | -0.121475 | -0.004038 | -0.059894 | 0.093045 | 0.386784 | -0.257748 | -0.577537 | 1.00000 |
| ca | 0.101389 | 0.070511 | 0.137979 | -0.072042 | -0.213177 | 0.115739 | 0.222682 | -0.08015 |
| thal | 0.062210 | 0.098803 | -0.032019 | -0.011981 | -0.096439 | 0.206754 | 0.210244 | -0.10476 |
| target | -0.144931 | -0.085239 | -0.028046 | 0.137230 | 0.421741 | -0.436757 | -0.430696 | 0.34587 |

In [47]:
```
df = cn[cn>=.40]
plt.figure(figsize=(6,5))
sns.heatmap(df, cmap="Blues")
```

Out[47]: <AxesSubplot:>




In [46]:
```
#correlation
cmap = sns.diverging_palette(260,-10, s=50, l=75, n=6, as_cmap=True)
plt.subplots(figsize=(10,10))
sns.heatmap(cn,cmap=cmap,annot=True,square=True)
plt.show()
```

In [48]:
```
df_new["choltrestbps"]=df_new["chol"]*df_new["trestbps"]
df_new
```

Out[48]:

| | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target | choltrestb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 | 33 |
| **1** | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 | 32 |
| **2** | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 | 26 |
| **3** | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 | 28 |
| **4** | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 | 42 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **298** | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 | 33 |
| **299** | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 | 29 |
| **300** | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 | 27 |
| **301** | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 | 17 |
| **302** | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 | 30 |

303 rows × 12 columns

In [52]:
```
x = df_new.drop("target", axis = 1)
y = df_new["target"]
```

In [62]:
```
from imblearn.over_sampling import ADASYN
adasyn = ADASYN(random_state=42)
x,y = adasyn.fit_resample(x,y)
```

In [54]:
```
len(x)
```

Out[54]: 303

In [55]:
```
#splitting data for training and testing
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest=train_test_split(x,y,test_size=0.25, random_st
```

In [56]:
```
#fitting training data to the model
from sklearn.linear_model import LogisticRegression
lr_model=LogisticRegression(random_state=0)
lr_model.fit(xtrain, ytrain)
```

Out[56]: LogisticRegression(random_state=0)

In [57]:
```
#predicting result using testing data
y_pred = lr_model.predict(xtest)
y_pred
```

Out[57]:
```
array([0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1,
       0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0,
       0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 0, 0, 1], dtype=int64)
```

```
In [58]: #model accuracy
         from sklearn.metrics import classification_report, accuracy_score,f1_score
         lr_cr=classification_report(ytest,y_pred)
         print(lr_cr)
```

```
              precision    recall  f1-score   support

           0       0.81      0.67      0.73        33
           1       0.78      0.88      0.83        43

    accuracy                           0.79        76
   macro avg       0.80      0.78      0.78        76
weighted avg       0.79      0.79      0.79        76
```

```
In [59]: #Decission Tree
         from sklearn.tree import DecisionTreeClassifier
         dt_model = DecisionTreeClassifier(criterion = "entropy", random_state=0)
         dt_model.fit(xtrain,ytrain)

Out[59]: DecisionTreeClassifier(criterion='entropy', random_state=0)
```

```
In [60]: y_dt_pred = dt_model.predict(xtest)
         y_dt_pred

Out[60]: array([0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1,
                0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0,
                0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1,
                0, 0, 1, 1, 1, 1, 1, 0, 0, 1], dtype=int64)
```

```
In [61]: #model accuracy
         dt_cr = classification_report(ytest,y_dt_pred)
         print(dt_cr)
```

```
              precision    recall  f1-score   support

           0       0.69      0.73      0.71        33
           1       0.78      0.74      0.76        43

    accuracy                           0.74        76
   macro avg       0.73      0.74      0.73        76
weighted avg       0.74      0.74      0.74        76
```

# 6. ADVANTAGE AND DISADVANTAGE

## 6.1 ADVANTAGE

**Accurate Predictions:** The project utilizes advanced machine learning algorithms, ensuring precise predictions of heart disease, aiding in timely interventions and personalized patient care.

**Real-time Predictions:** The system provides real-time predictions, enabling healthcare professionals to make swift decisions based on the latest patient data.

**User-Friendly Interface:** The user interface is intuitive, allowing easy input of patient data and interpretation of results, making it accessible for healthcare practitioners.

**Ethical Considerations:** The project emphasizes data privacy, ensuring patient confidentiality, and addresses biases, promoting fairness and equity in healthcare predictions.

**Continuous Improvement:** Feedback mechanisms and regular updates ensure the system remains relevant, adapting to emerging medical knowledge and evolving patient demographics.

**Enhanced Decision Making:** Accurate predictions empower healthcare professionals to make informed decisions, leading to improved patient outcomes and a higher quality of healthcare delivery.

## 6.2 DISADVANTAGE

**Data Dependency:** The accuracy of predictions relies heavily on the quality and representativeness of the training data. Biased or incomplete data can lead to skewed predictions.

**Model Complexity:** Advanced machine learning models might be complex, making it challenging to interpret and explain predictions, especially in critical healthcare decisions where interpretability is crucial.

**Resource Intensive:** Developing and maintaining the system, especially in real-time applications, might require significant computational resources and skilled personnel, making it resource-intensive.

**Over-Reliance on Technology:** There's a risk of over-reliance on the technology, potentially leading to reduced human judgment and interaction in healthcare decisions, which might not always be ideal.

**Regulatory Challenges:** Adhering to healthcare regulations and standards, especially concerning data privacy and ethical considerations, can present challenges, requiring continuous monitoring and compliance.

**User Adoption:** Healthcare professionals might require training and adaptation to effectively use the system, and initial resistance to technology adoption could be a hurdle.

# 7. CONCULSION

In the realm of healthcare, accurate prediction of heart disease is paramount for timely interventions and improved patient outcomes. This project embarked on a comprehensive journey, leveraging advanced machine learning techniques to develop a robust and efficient heart disease prediction system.

Through meticulous data preprocessing, including cleansing and feature engineering, the dataset was refined, ensuring the integrity and relevance of the information used for predictions. Employing state-of-the-art algorithms such as Random Forest and Gradient Boosting, the system achieved exceptional accuracy, enabling precise identification of potential cardiac risks in patients.

The system’s user-friendly interface facilitates seamless interaction, allowing healthcare professionals to input patient data and receive real-time, reliable predictions. Ethical considerations, including data privacy and bias mitigation, were embedded into the system’s core, ensuring fairness and confidentiality in its operations.

The success of this project lies not only in its accuracy but also in its adaptability. Continuous improvement mechanisms, such as feedback loops and model updates, guarantee that the system evolves alongside medical advancements, maintaining its relevance and efficacy over time.

In conclusion, this heart disease prediction system stands as a testament to the potential of machine learning in revolutionizing healthcare. By combining precision, efficiency, and ethical considerations, this project contributes significantly to the field of cardiac health diagnostics, empowering healthcare professionals with a powerful tool for personalized patient care and ultimately enhancing the quality of healthcare delivery.

# 8. REFERENCE


❖https://towardsdatascience.com/heart-disease-uci-diagnosis-prediction b1943ee835a7

❖https://www.frontiersin.org/articles/10.3389/fmed.2023.1150933/full

❖https://towardsdatascience.com/heart-disease-uci-diagnosis-prediction b1943ee835a7#70b