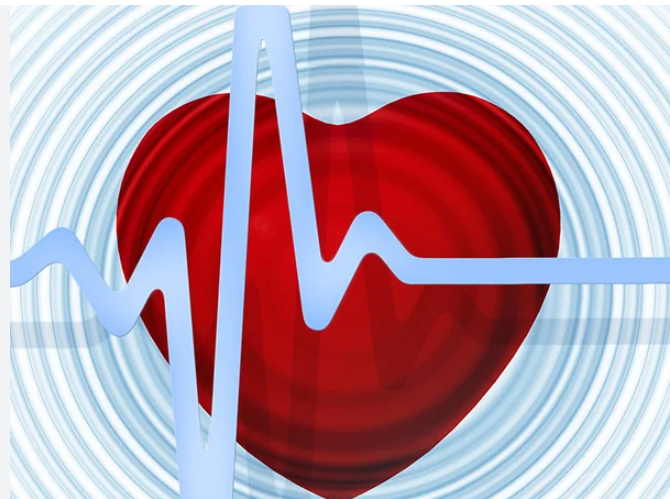
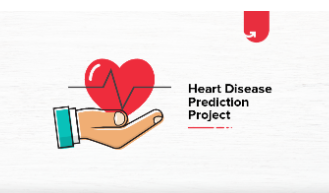
**Heart Attack Prediction Model**

Heart disease, a prevalent global health concern, can benefit from data-driven insights. Leveraging machine learning, we aim to predict heart disease occurrence by analyzing medical histories, physiological measurements, and lifestyle factors. Through model training and evaluation, our study seeks to accurately differentiate patients with and without heart disease. Challenges include data quality and model complexity, demanding iterative refinement. This research aligns with the growing need for early detection, intervention, and personalized treatment strategies, ultimately improving cardiovascular healthcare outcomes.

[Features Information:](https://archive.ics.uci.edu/ml/datasets/iris)

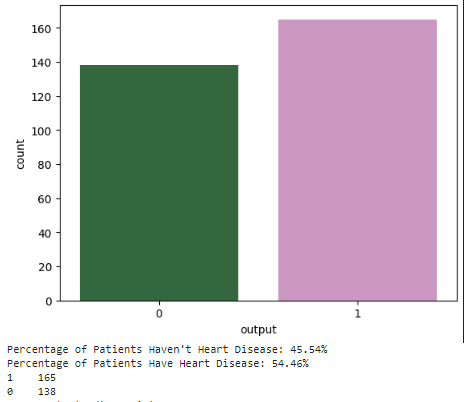
|  |  |  |
| --- | --- | --- |
| **Index** | **Feature** | **Description** |
| **1** | **age** |  |
| **2** | **sex** |  |
| **3** | **cp** | **Chest Pain** |
| **4** | **trtbps** |  |
| **5** | **chol** | **Cholesterol** |
| **6** | **fbs** | **fasting blood sugar** |
| **7** | **restecg** | **Resting electrocardiographic measurement** |
| **8** | **thalachh** | **The person’s maximum heart rate achieved** |
| **9** | **exng** | **exercise induced angina** |
| **10** | **oldpeak** | **The person’s maximum heart rate achieved. Exang: Exercise induced angina** |
| **11** | **slp** | **Speech and Language Disorders** |
| **12** | **caa** | **A coronary artery anomaly** |
| **13** | **thall** | **thalassemia** |
| **14** | **output** | **Target** |

|  |
| --- |
|  |
|  |

# ****Data Visualization****

# 

Total distribution of output is 303 (165 has heart disease and 138 are not)



# ****Gender distribution for the study****

# 

# 

The study distribution shows that male is almost 2/3 as compared to 1/3 of female

# ****Correlation between sex and Heart Disease****

# 

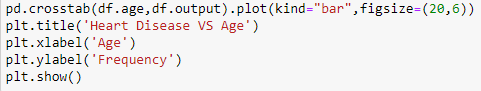
# 

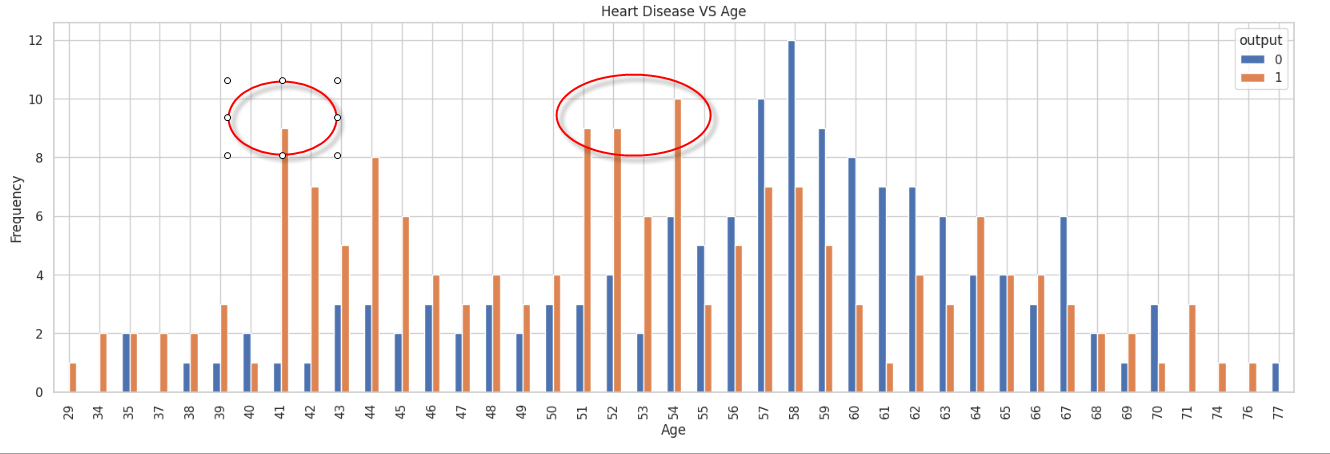
# Out of 96 Female, 72 were diagnosed with no disease while 24 have disease.

# Out of 207 Male, 93 were diagnosed with no disease while 114 have disease

# The percentage for heart disease for female is 25% (24/96) while it jumps to 45% for male (114/207)

**Heart Disease VS Age**





From the graph above, we can see that most of the heart disease samples are at age 41 as well as 51,52 and 54.

# ****SLP and output****

# 

# There are high number of positive samples if SLP (Speech and Language disorders at stage 2)

# ****Correlation between FBS and output****

# 

# 

# The above graph shows that percentage of heart disease increase when it is > 120mg/dl while the percentage almost same when it is <120mg/dl

# ****Correlation between Chest pain type and output****

# 

# 

# Graph shows chest pain type 2 has high potential for heart disease while it is minimal for type 3

# ****correlation coefficients (Features VS Output)****

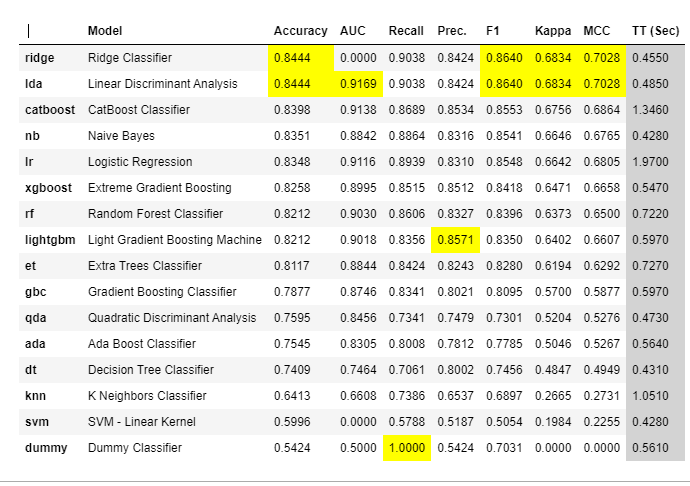
# 

# 

From the above graph, we observe that CP (Chest Pain)/ thalachh/Exng(exercise induced angina) and old peak are the highest features that correlate with the output.

**Metrics comparison for Different Machine Learning Models using Pycaret Library**





In the context of detecting heart disease or any medical condition, minimizing false negatives (cases where the model fails to identify a positive case) is crucial. Therefore, prioritizing recall (sensitivity) as a metric makes sense, as it measures the ability of the model to correctly identify most of the positive cases. However, it's important to consider a balanced approach that takes into account precision, recall, and possibly the F1-score, as maximizing recall could lead to increased false positives. Striking a balance between these metrics ensures a comprehensive evaluation of the model's performance in detecting heart disease so Ridge Classifier and Linear Discriminant Analysis are best models to consider in this analysis