**Health care: Heart attack possibility in Cleveland**

**Introduction:**

A heart attack happens every 40 seconds in the United States, making it a leading cause of death among patients. More than 805,000 Americans have a heart attack every year, and in most cases, patients suffer from recurrent heart attacks. Therefore, it is important to understand the leading factors of a heart attack in order to detect and prevent heart attacks.

**Data Source:**

The data source is from Cleveland area hospitals, available on Kaggle, and the direct link is included below. The dataset consists of 14 variables that can be analyzed for the presence of heart disease in patients. This data set is interesting because we analyze these variables and determine which ones are good indicators for predicting heart attack.

Link: <https://www.kaggle.com/nareshbhat/health-care-data-set-on-heart-attack-possibility>

**Model and Analysis:**

A model was created which included thirteen variables and the target. A list of all fourteen variables is given below in table 1. A logic regression analysis was performed to determine which variables are the best indicators to predict a heart attack. A total of two runs were completed. The model was evaluated using the pseudo r-squared, statistical accuracy, MCC, precision, recall, specificity, f-score, confusion matrix, Receiver Operating Characteristic (ROC) and Area Under Curve (AUC). This model was also then used to history match and predict number of heart attacks in patients check for reliability of the model again.

Furthermore, to better understand the variables and their relationship to the target, data visualization analysis was also completed on each variable.

*Table 1: List of all variables included in the data set and their description*

|  |  |
| --- | --- |
| **Field Name** | **Medical Name and description** |
| Target (0 = unlikely or 1 = likely) | Heart Disease - likelihood of patient experiencing a heart attack |
| Age | Age of patient |
| Sex | Sex of patient |
| Cp (Value 0: typical angina, Value 1: atypical angina) | Chest pain type |
| Trestbps (mm Hg on admission to the hospital) | Patients resting blood pressure |
| Chol (mg/dl) | Serum cholesterol |
| Fbs (more than 120mg/dl) | Fasting blood sugar |
| Restecg (0 = normal, 1 = having ST-T wave abnormality) | Resting electrocardiographic results of patient |
| thalach | Maximum heart rate achieved |
| Exang (1 = yes; 0 = no) | Exercise induced angina |
| oldpeak | ST depression induced by exercise relative to rest ('ST' relates to positions on the ECG plot) |
| Slope | The slope of the peak exercise ST segment |
| CA (0-3) | Number of major vessels coloured by fluoroscopy |
| Thal (0 = normal; 1 = fixed defect; 2 = reversible defect) | A blood disorder called thalassemia |

**Conclusion:**

From the logic regression model and the analysis, it was found that of the thirteen variables, seven variables are a strong predictor for heart attacks. Those variables are sex, chest pain, exercise-induced angina (exang), the number of major vessels (ca), thalassemia (thal), maximum heart rate (thalach) and the slope of the peak exercise ST segment (slope). The model predicted with 87% accuracy. The model was also determined to have a highly, almost balanced sensitivity and specificity which was another assurance that the model is highly reliable. The Area under the ROC curve is 93%, which is very satisfactory.

Additional quick facts discovered upon data analysis:

* The classifier made a total of 293 predictions (i.e., 293 patients for the heart attack). Out of those 293 patients, the classifier predicted that 163 patients are susceptible to heart attack, and 130 patients are not. In reality, 162 patients are susceptible to heart attack, and 131 patients are not. These are another indicator the model is reliable.
* It was found that a heart rate higher than 140 bpm shows a higher possibility of heart attack.
* It was determined that patients with resting blood pressure among 120 mm Hg to 150 mm Hg have a higher risk of getting a heart attack.