PRCP-1016-HeartDieseasePred

Problem Statement

Task 1:-Prepare a complete data analysis report on the given data.

Task 2:- Create a model predicting potential Heart Diseases in people using Machine Learning algorithms.

Task3:-Suggestions to the Hospital to awake the predictions of heart diseases prevent life threats.

Dataset Link:

- Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Four out of 5CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age. Heart failure is a common event caused by CVDs and this dataset contains 11 features that can be used to predict a possible heart disease.
- People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help.

Domain: Healthcare

Link : https://d3ilbtxij3aepc.cloudfront.net/projects/CDS-Capstone-Projects/PRCP-1016-HeartDieseasePred.zip

Dataset

1. There are 14 columns in the dataset, where the patient_id column is a unique and random identifier. The remaining 13 features are described in the section below.

2. · slope_of_peak_exercise_st_segment (type: int): the slope of the peak exercise ST segment, an electrocardiography read out indicating quality of blood flow to the heart

3. · thal (type: categorical): results of thallium stress test measuring blood flow to the heart, with possible values normal, fixed_defect, reversible_defect

4. · resting_blood_pressure (type: int): resting blood pressure

5. · chest_pain_type (type: int): chest pain type (4 values)

6. · num_major_vessels (type: int): number of major vessels (0-3) colored by flourosopy

7. · fasting_blood_sugar_gt_120_mg_per_dl (type: binary): fasting blood sugar > 120 mg/dl

8. · resting_ekg_results (type: int): resting electrocardiographic results (values 0,1,2)

9. · serum_cholesterol_mg_per_dl (type: int): serum cholestoral in mg/dl

10. · oldpeak_eq_st_depression (type: float): oldpeak = ST depression induced by exercise relative to rest, a measure of abnormality in electrocardiograms

11. · sex (type: binary): 0: female, 1: male

12. · age (type: int): age in years

13. · max_heart_rate_achieved (type: int): maximum heart rate achieved (beats per minute)

14. · exercise_induced_angina (type: binary): exercise-induced chest pain (0: False, 1: True)

Model Comparison Report

Create a report stating the performance of multiple models on this data and suggest the best model for production.

Report on Challenges faced

Create a report which should include challenges you faced on data and what technique used with proper reason.

Note:-All above task has been created on single jupyter notebook and share the same while final submission of project.