Cardiovascular Disease Prediction Analysis

Coursera Capstone

by

Chris Buys

July, 2021

1. Introduction

1.1 Background

Cardiovascular disease is one of the leading causes of death globally and is the leading cause in the United States. Heat disease costs the United States approximately 219 billion dollars each year which includes the cost of health care services, medicines and lost of productivity due to death (Centers for Disease Control and Prevention, 2020).

1.2 Problem

If people with or without cardiovascular disease could be identified with a certain amount of certainty resources could be better planned for and the target group could be treated more effectively. Especially if we can eliminate false positives (outliers) in treating cardiovascular disease.

1.3 Interest

Addressing this problem would yield benefits for physicians as they get to focus on treating the right people and government who can better plan and allocate resources towards the treatment of cardiovascular disease.

2. Data

2.1 Data Description

The data set consists out of twelve total features which are categorised into three types of input features and one target feature (kaggle.com, n.d.). The input features include objective information (factual), examination (result of medical examination) and subjective information (information given by the patient). The table below describes the features of the data set:

Feature	Category of Feature	Metric	Unit of Measurement
Age	Objective	age	int(days)
Height	Objective	height	int(cm)
Weight	Objective	weight	float(kg)
Gender	Objective	gender	categorical code
Systolic Blood Pressure	Examination	ap_hi	int
Diastolic Blood Pressure	Examination	ap_lo	int
Cholesterol	Examination	cholestorol	1: normal, 2: above normal, 3: well above normal
Glucose	Examination	gluc	1: normal, 2: above normal, 3: well above normal
Smoking	Subjective	smoke	binary
Alcohol Intake	Subjective	alco	binary
Physical Activity	Subjetive	active	binary
Presence or Absence of cardiovascular disease	Target Variable	cardio	binary

2.2 Feature Selection

All Features are applicable to include in this analysis however, it would be informative to split the data into features containing the subjective features and not containing to indicate how truthful patients are about this information. Often people either over or under exaggerate and this can introduce issues when trying to treat patients. Depending on the outcome this could give an indication if whether or not a patients subjective information is needed during this type of analysis.

3. Methodology

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

4. Results

Results section where you discuss the results.

5. Discussion

Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

6. Conclusion

Conclude the report

7. References

Centers for Disease Control and Prevention (2020). *Heart disease facts & statistics*. [online] Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/heartdisease/facts.htm.

kaggle.com. (n.d.). *Cardiovascular Disease dataset*. [online] Available at: https://www.kaggle.com/sulianova/cardiovascular-disease-dataset.