



METIS



SDAIA

الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

Proposal:

Predicting Cardiovascular Disease with Classification Machine Learning Algorithms

- Lama Ibrahim Alharbi

Table of Contents

1 INTRODUCTION	3
1.1 BACKGROUND.....	3
1.2 MOTIVATION AND IMPORTANCE	3
2 PROBLEM STATEMENT	4
3 RESEARCH METHODOLOGY.....	4
3.1 DATA SOURCE.....	4
3.2 RESEARCH QUESTION	5
3.3 LIBRARIES AND ALGORITHM	5
REFERENCE.....	6

1 Introduction

1.1 Background

Heart disease affects millions of people, in medicine falls under the category of cardiovascular disease, which encapsulates various conditions that affect the heart. Cardiovascular disease is the leading cause of death worldwide over the past decades. It relates the many risk factors for cardiovascular disease to the need for time to obtain accurate, reliable and reasonable methods for early diagnosis to achieve rapid disease management [1]. Data mining is a commonly used technique for processing big data in healthcare. Researchers apply various data mining and machine learning techniques to analyze huge complex medical data, helping healthcare professionals to predict heart disease. A review of the empirical literature over the past 20 years addresses compliance across four cardiovascular risk reduction regimens: drug therapy, exercise, nutrition, and smoking cessation [2]. In this project I will go through the idea of developing machine learning model to predict whether the patient has potential of cardiovascular disease or not.

1.2 Motivation and importance

Fast increasing of heart disease across the world suggests that currently used primary cardiovascular disease prevention strategies are not sufficiently effective [3]. Science has demonstrated the cardiovascular benefits of regular physical activity and healthy food choices. Furthermore, a lot of pharmacotherapies are now available to decrease the occurrence of cardiovascular disease in symptomatic patients, also among asymptomatic apparently healthy persons at high risk. To reduce these incidents and save patients this project aims to spot the light on the importance of this disease and predict the patient health condition from his/her history.

2 Problem statement

Predict presence or absence of cardiovascular disease using the patient examination results by building classification machine learning models to classify the patients to be healthy or suffering from cardiovascular disease based on the given attributes.

3 Research methodology

This project aims to predict the presence of heart disease in the patient to develop and discover the probability of having heart attack which is useful in the medical field for clinicians and patients. To accomplish the aim, later I will discuss the use of various machine learning algorithms on the data set and dataset analysis. This project will additionally depict which attributes contribute more than the others to anticipation of higher precision. The tool will be used is Jupyter Notebook for developing the model using Python.

3.1 Data source

For this project, I have used dataset [Kaggle](#). It comprises a real dataset that were collected at the moment of medical examination of +7000 examples of data with 13 various attributes (12 predictors; 1 target) below table 1 describes the features in details.

There are 3 types of input features:

- Objective: factual information;
- Examination: results of medical examination;
- Subjective: information given by the patient.

Table 1 Dataset description

Feature	Feature type	Column name	Feature value type
Age	Objective	age	Int(days)
Height	Objective	height	Int(cm)
Weight	Objective	weight	Float(kg)
Gender	Objective	gender	Object(categorical code)
Systolic blood pressure	Examination	ap_hi	int
Diastolic blood pressure	Examination	ap_lo	int
Cholesterol	Examination	cholesterol	Int (ratio: 1: normal, 2: above normal, 3: well above normal)
Glucose	Examination	gluc	Int (ratio: 1: normal, 2: above normal, 3: well above normal)
Smoking	Subjective	smoke	Int(binary)
Alcohol intake	Subjective	alc	Int(binary)
Physical activity	Subjective	active	Int(binary)
Presence or absence of cardiovascular disease	Target	cardio	Int(binary)

3.2 Research Question

1. Is there probability of someone whose male and +60 with 85 Kg and diastolic blood pressure equals 100 and systolic blood pressure over 130 to suffer from cardiovascular disease based on the given attributes?

3.3 Libraries and Algorithm

The libraries I will use are Pandas for dealing with data, Sklearn for predictive models, Matplotlib and Seaborn for visualization. The problem I'm focusing on is classification problem. I will use different algorithms such as Logistic regression, Decision tree, Random forest, KNN, etc. Then I will compare the results to see which model give the highest accurate results.

Reference

[1] (Devansh Shah, Heart Disease Prediction using Machine Learning Techniques, 2020)

[2] (Lora E. Burke, Compliance with cardiovascular disease prevention strategies: A review of the research , 1997)

[3] (Valery L Feigin¹, Primary prevention of cardiovascular disease through population-wide motivational strategies: insights from using smartphones in stroke prevention, 2017)