

AI for Everyday Life: Investigating Factors Affecting Heart Attack using Linear Regression

CS 180 Project Proposal

Group Members:

Chase Gonzaga (WFW)

Althea Manansala (WFY)

Jeff Pecson (WFW)

Brill Riña (WFY)

I. Project Summary

In 2022, PSA released the causes of deaths in the Philippines. Data shows that heart diseases are the leading causes of mortality in the Philippines. This study aims to analyze the relationship between certain factors (e.g. age, cholesterol, and sugar levels) that influence heart attack in the Philippines' setting using Linear Regression.

II. Background

Heart attacks, also known as a myocardial infarction, is a major cause of mortality in the Philippines. This occurs when there is not enough blood supplied to a certain part of the heart muscle due to a restriction or blockage which can lead to tissue damage or death (Cleveland Clinic, 2022). Cardiovascular diseases, ischemic heart disease in particular, was the leading the causes of mortality in the Philippines in 2022, accounting for 18.5% of the registered deaths (Cabico, 2023). Heart disease can be influenced by the changing lifestyle patterns in the country and risk factors such as obesity, blood pressure, and diabetes. They are a common cause of hospitalization in the country with an in-hospital fatality rate of 8.2% (Reyes et al., 2020).

This project aims to predict the risk of heart attack for an individual given their vital signs. This may help in the prevention of heart attacks by alerting and furthermore, it might provide more insight on the relation of different vital signs and how they may potentially contribute to heightening risk of heart attacks. Professionals in the medical field may also find some use for this by finding potential early warning signs of heart failure so that preventative treatment can begin early for those at risk.

III. Materials and Methods

We will use a **linear regression model** to find a linear relationship between the occurrence of heart attacks and other factors such as age, sex, chest pain type, etc. To train the model, we will be using the labeled dataset **Heart Attack Analysis & Prediction Dataset**. This dataset contains the following labels:

- **age** - age of the patient
- **sex** - sex of the patient
- **cp** - chest pain type
- **trtbps** - resting blood pressure (in mm Hg)
- **chol** - cholestoral in mg/dl fetched via BMI sensor

- **fbs** - fasting blood sugar
- **restecg** - resting electrocardiographic results
- **thalachh** - maximum heart rate achieved
- **oldpeak** - previous peak
- **slp** - slope
- **caa** - number of major vessels
- **thall** - thalium stress test result
- **exng** - exercise induced angina
- **output** - target variable (the dependent variable)

We import the linear regression model from the **scikit learn** library and use **pandas** library to read the dataset. After we have fit the model to our training data, we can use R2 score also from the scikit learn library to determine the accuracy of the model. To visualize the model, we can plot the regression line using **matplotlib** to get a better visual understanding of the model.

IV. References

Heart Attack Analysis & Prediction Dataset

<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>

Cabico, G. K. (2023, January 25). **Covid-19 not among top 10 causes of death in 2022**. Philstar.com.

<https://www.philstar.com/headlines/2023/01/25/2240134/covid-19-not-among-top-10-causes-death-2022>

Reyes, E. B., Punzalan, F. E., Reganit, P. F., & Agbayani, M.-J. F. (2020). **Epidemiology, treatment patterns and in-hospital outcomes of patients with heart failure with reduced ejection fraction: An analysis of the heart failure registry of the Philippine Heart Association.**

ASEAN Heart Journal, 27(1), 1-5. <https://doi.org/10.31762/ahj2027.0101>

Cleveland Clinic. (2022, October 30). **Heart attack: Symptoms and treatment.**

<https://my.clevelandclinic.org/health/diseases/16818-heart-attack-myocardial-infarction>