

Health Risk Analysis & Disease Prediction Report

1. Introduction

This project focuses on analyzing health and lifestyle data to understand factors contributing to heart disease. Exploratory Data Analysis (EDA) and Logistic Regression were applied to extract insights and build a predictive model.

2. Dataset Description

The dataset consists of patient health indicators such as age, BMI, blood pressure, cholesterol, glucose level, smoking habits, and physical activity. The target variable indicates the presence or absence of heart disease.

3. Assumptions

- Data represents adult patients
- No missing values present
- Observations are independent
- Linear relationship assumed for regression modeling

4. Exploratory Data Analysis

EDA revealed that higher BMI, blood pressure, cholesterol, and glucose levels are strongly associated with heart disease. Smokers and individuals with low physical activity levels showed higher risk.

5. Visual Analysis

Visualizations included distribution plots, correlation heatmaps, and categorical comparisons. These visuals helped identify patterns and relationships among variables.

6. Regression Modeling

A Logistic Regression model was used due to the binary nature of the target variable. The model identified age, BMI, blood pressure, smoking status, and physical activity as significant predictors of heart disease.

7. Conclusion

The analysis confirms that lifestyle and physiological factors play a crucial role in heart disease risk. This project demonstrates the practical use of data analysis and machine learning in healthcare decision-making.

Author: Toliha Abiodun Ahmad