

HEART DISEASE PREDICTION USING MACHINE LEARNING AND DATA MINING TECHNIQUES

Biyanka De Silva

A dissertation submitted in partial fulfilment of the requirement for
Master of Science degree in Business Analytics

Department of Business School
Informatics Institute of Technology, Sri Lanka
in collaboration with
Robert Gordon University, Aberdeen

2022

Abstract

In this research project, the overall research and the project implementation is focused on predict a heart failure or a disease based on Machine Learning, Data Mining Techniques make use of particular dataset. The heart is an essential part of the human body. It pumps blood to every single part of the human body. If it stops to work properly, then the brain and numerous other organs will stop working, and within a few minutes, a person will die. Lifestyle changes, work-related stress and poor eating habits contribute to an increase in the rate of several heart-related diseases.

The goal of the study is to develop a model that can help clinicians make better medical judgments using computerized and automated patient information. And also, will be discussing and evaluating about different hypotheses that can impact on the heart diseases. This model will reduce medical mistakes, increase patient safety, eliminate undesired practice variation, and improve patient outcomes. This model can facilitate to enhance the quality of clinical decisions by using machine learning and artificial intelligence algorithms.

At this point the scope of the research project is that combining support for medical decisions with computerized patient histories can decrease health check mistakes, improve patient protection, reduce unnecessary procedure variability, and enhance patient outcome. This research is promising as a model for modeling and analyzing data, e.g., data mining, has the potential to generate a wealth of information that can help significantly improve the value of medical outcomes.

This study will employ an exploratory research approach. It is an approach of quantitative research. This project contains, Exploratory Data Analysis, Data Preprocessing, Modeling Phase, Model Evaluation and Prediction System – Web Application. Python, Jupyter Notebook (anaconda3), PyCharm IDE, HTML, REST API and Flask web allocation framework technologies used for developed this research project.

Extra Tree Classifier, K-Nearest Neighbors, Random Forest Classifier, Gradient Boosting Classifier, Logistic Regression, Support Vector Machine, AdaBoost Classifier, Gaussian Naive Bayes and Decision Tree Classifier machine learning

technique used for this project and final evaluation done by using Accuracy, Precision, and Recall, F1 Score, AUC ROC, Confusion Matrix