# CIND 820 – Capstone Project Abstract: Prediction of Coronary Artery Disease

By: Sherriza Khan (501145108)

# Abstract

Heart disease is second most common cause of death in Canadians and the primary cause of hospitalisations in Canada (Government of Canada, 2017). Each year roughly 92,000 Canadians aged 40 and above are diagnosed with heart disease (Government of Canada, 2017). The phrase heart disease broadly relates to a variety of heart conditions. Coronary heart disease also referred to as ischemic heart disease is the most frequently diagnosed form of heart disease (Government of Canada, 2017). The condition arises from the development of atherosclerotic plaques within the walls of arteries. The presence of the plaques reduces blood flow within arteries and promotes the formation of blood clots that may lead to myocardial infarction (Karp, 2010). Early detection of coronary heart disease as well as identifying primary risk factors may reduce the incidence of fatal or severe heart disease and improve patient quality of life.

The objectives of the project are to develop a classification model of coronary heart disease and identify which risk factor variables contribute most acutely to the development of coronary heart disease. In addition, this project seeks to determine if there are differences in risk factor variables between males and females. The UCI Machine Learning Repository Heart Disease dataset will be used for the purposes of this research project[[1]](#footnote-1). Datasets collected from 4 locations will be aggregated for a total of 920 records and 11 common features including but not limited to age, sex and serum cholesterol levels (UC Irvine Machine Learning Repository, 2023). The patient records were collected from Cleveland with 303 records, Hungary with 294 records, Long Beach with 200 records and Switzerland with 123 records. The aggregated dataset contains 6 categorical variables and 5 integer variables and was collected from patients undergoing angiography at the above-mentioned locations in 1988 (Detrano, et al., 1989).

This project will use Python to preprocess the data and develop the classifiers. Exploratory data analysis will be performed using the Pandas Profiling library. Gaussian Naïve Bayes, Logistic Regression and Decision Tree classifiers will be developed and compared for accuracy and precision. The scikit-learn library will be used with the classes GaussianNB, LogisticRegression and DecisionTreeClassifier to develop the models (scikit-learn developers, 2023). Accuracy, precision and confusion matrices will be generated and compared for each model using accuracy\_score, precision\_score and confusion\_matrix classes within the sklearn.metrics module. Forward feature selection will be applied prior to logistic regression and naïve bayes classification to identify those variables with the greatest impact on heart disease risk.

# Bibliography

Detrano, R., Janosi, A., Steinbrunn , W., Pfisterer, M., Schmid, J.-J., Sandhu, S., . . . Froelicher, V. (1989, August). International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease. *Cardiology, 65*(5).

Government of Canada. (2017, July 18). *Heart disease in Canada: Highlights from the Canadian Chronic Disease Surveillance System*. Retrieved from Government of Canada: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/heart-disease-canada-fact-sheet.html

Karp, G. (2010). *Cell and Molecular Biology Concepts and Experiments* (6th Edition ed.). Hoboken: John Wiley & Sons, Inc.

scikit-learn developers. (2023, 9 22). *1. Supervised learning*. Retrieved from scikit learn: https://scikit-learn.org/stable/supervised\_learning.html#supervised-learning

UC Irvine Machine Learning Repository. (2023, 09 24). *Heart Disease*. Retrieved from https://archive.ics.uci.edu/dataset/45/heart+disease: https://archive.ics.uci.edu/dataset/45/heart+disease

1. The Heart Disease dataset is available through the UCI Machine Learning Repository at the link https://archive.ics.uci.edu/dataset/45/heart+disease [↑](#footnote-ref-1)