**Abstract**

**Background.** Cardiovascular disease is the leading cause of death worldwide. Although modern technologies provide efficient and accurate diagnosis of heart disease, identifying people at-risk would enable early prevention and treatment, which is often preferable than the previous. Many machine learning techniques is being used to serve this purpose with an average accuracy of 80%, which, although is a great advancement, still needs improvement.

**Aim.** Our goal is to predict the risk of developing heart disease, with an accuracy rate over 80%, based on some easily obtainable medical records such as age, gender, BMI, blood glucose, cholesterol, levels of physical activity, alcohol consumption, smoking habit, etc.

**Methodology.** Among numerous models of risk prediction, we chose to use Naïve Bayes and Decision Tree along with standard statistical tests like ANOVA, Chi-Square, T-Test, etc, for selecting best attributes.

**Results.**

**Conclusion.**

Example format

<https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-019-0681-4>

==================================================================================

Written by Jing Wen

**Background**

With the increase and successful implementation of machine learning in various predictive tasks in the real world, attracting the interest in the medical field in applying the same technique on available data. Cardiovascular disease being one of the leading causes of death worldwide, although modern technologies provide accurate diagnosis of cardiovascular disease, most often diagnosis takes too much time, or it is too late. Since identifying people at-risk would enable early prevention and treatment, which is often preferable than the previous. By using openly available software and public domain data, machine learning techniques implementation and evaluation will be done to serve this purpose.

**Aim (Mostly similar)**

Our goal is to create a predictive model that can predict the risk of developing cardiovascular disease in patients, with an accuracy rate over 80%, based on some easily obtainable medical records such as age, gender, BMI, blood glucose, cholesterol, levels of physical activity, alcohol consumption, smoking habit, etc.

**Methods**

Demonstration on the usage of machine learning algorithms in building predictive models for cardiovascular diseases diagnosis using descriptions of data records. The algorithms chosen are Naïve Bayes and Decision Tree along with standard statistical test XXX in selecting the best attributes. The dataset used is obtained from a publicly available source, Kaggle will be split randomly into training and testing samples.

Algorithms are trained using the data from the training sample before using the test sample to predict the target where identification of presence or absence of cardiovascular disease in patients are done. Performance of the predictive models is completed using matrices such as accuracy, recall, precision and f1-score. The steps using in the algorithm development using open-source tools R will be provided in this paper.

**Results**

**Conclusion**