HEART ATTACK RISK PREDICTION (Methodology)

BY: Folowosele, Oluwadamilola Sarah [208507]

SUPERVISOR: Dr. Adejoke O. Olamiti

RECAP: PROBLEM STATEMENT

- Cardiovascular disease events could be prevented or delayed if patients are aware of the disease risk early enough. Existing data mining models have been used to predict the existence (i.e. presence or absence) of the disease rather than the risk of patients developing the disease. There is a need to determine the risk of a patient developing the disease in order to mitigate its incidence.

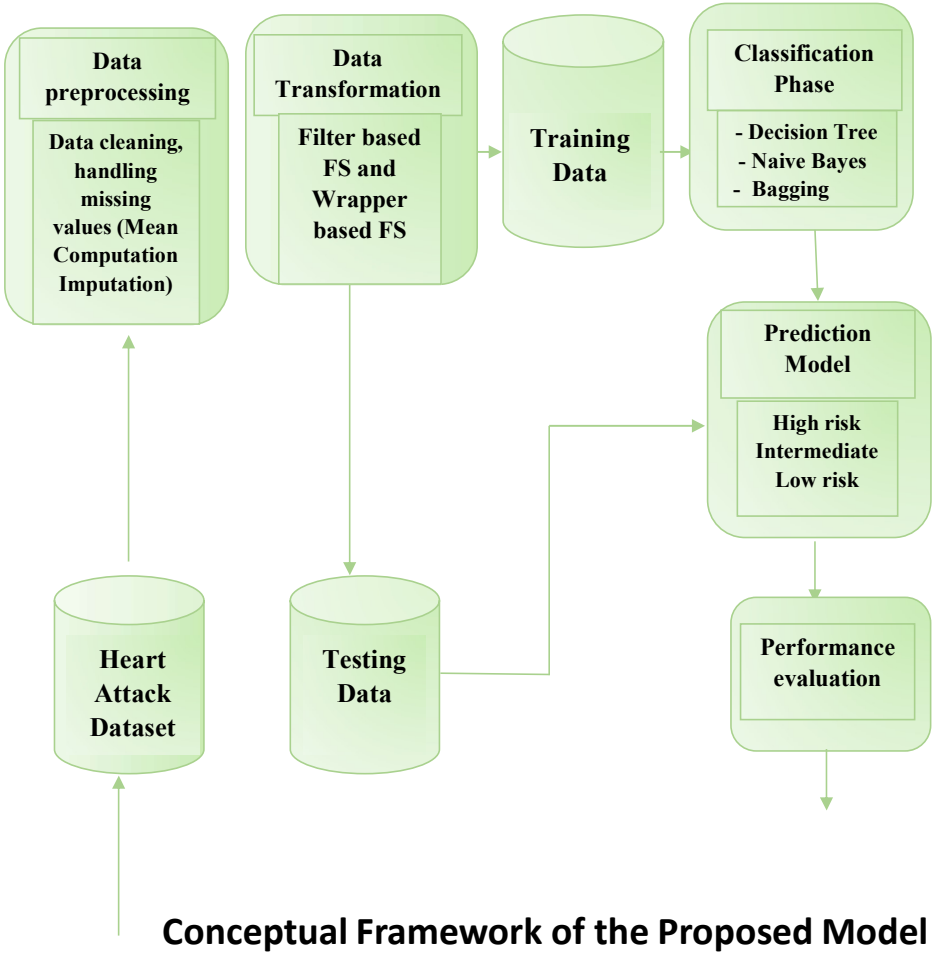
RECAP: AIM

- The aim of this research is to develop a model that will predict the risk of an individual developing heart attack.

RECAP: OBJECTIVES

- To formulate a predictive model to determine the likelihood of having heart attack in patient
- Simulate the model formulated using a collection of knowledge discovery and machine learning tools.
- Evaluate the performance of the model

METHODOLOGY



Conceptual Framework of the Proposed Model

DATA PRE-PROCESSING

- During the data collection phase, some of the collected data might have features that would not enhance the behavior of the model. These features will be processed to ensure optimal behavior of the model.

DATA TRANSFORMATION

- This process is used to determine which set of variables are deemed more relevant and effective for heart attack risk prediction, identifying the variables will help to improve the performance of the supervised machine learning algorithms' performance and also reduce the complexity of the model.

- The basic algorithm for implementing the hybrid feature selection algorithm used in this study is by performing the filter based feature selection using information gain ratio (which calculates the relevance of the attributes based on the information gain and split criteria, and assigns weights to them accordingly) and running the result using the wrapper based feature selection.

PREDICTION MODEL

The model formulation Will be done using the selected supervised machine learning algorithm of

- Naives Bayes (NB)
- Decision Tree (DT)
- Bagging algorithms

BAGGING ALGORITHM

Input: training set S, Inducer I, integer T (Number of bootstrap samples).

1: For i = 1 to T {

2: S' = bootstrap sample from S (i.i.d. sample with replacement).

3: $C_i = I(S')$

}

4: $C^*(x) = \underset{y \epsilon Y}{\text{argmax}} \sum_{i\ C_i(x)=} 1$ $(the\ most\ often\ predicted\ label\ y)$

//Where y is the class label and x is a tuple of the dataset D of the inducer.

Output: classifier C^*

TOOLS

- PYTHON

THANK YOU