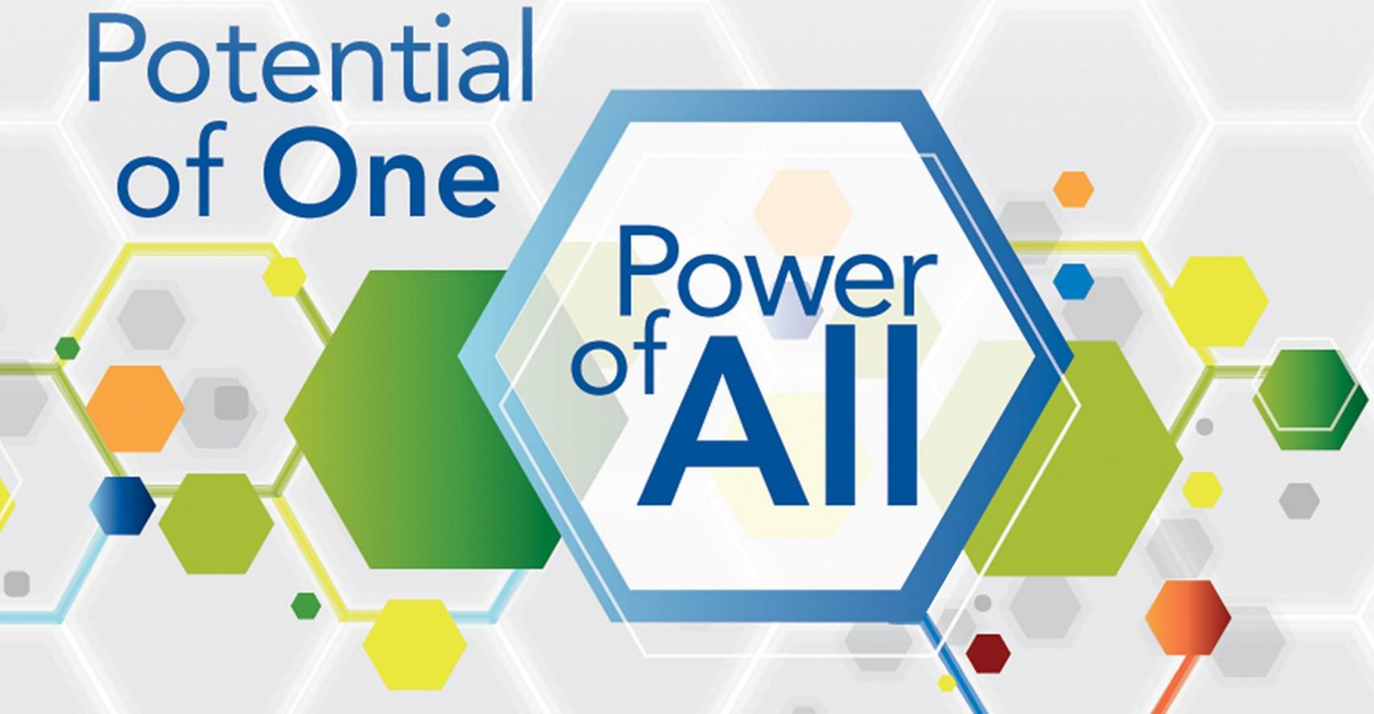
SAS GLOBAL FORUM

Survival of Your Heart: Analyzing the effect of stress on a cardiac event and predicting the Survival Chances

E-poster 1892



Survival of Your Heart: Analyzing the effect of stress on a cardiac event and predicting the Survival Chances

Yogananda Domlur Seetharama and Sai Vijay Kishore Movva

MSIS and SAS® & OSU Data Mining Certificate Program, Oklahoma State University.

Faculty Advisor: Dr. Goutam Chakraborty

Abstract

One in every four people dies of heart disease in the United States, and stress is an important factor which contributes towards a cardiac event. As the condition of the heart gradually worsens with age, the factors that lead to a myocardial infarction when the patients are subjected to stress are analyzed. The data used for this project were obtained from a survey conducted through the Department of Biostatistics at Vanderbilt University.

The objective of this poster is to predict the chance of survival of a patient after a cardiac event. Then, by using decision trees, neural networks, regression models, bootstrap decision trees, and ensemble models we predict the target which is modeled as a binary variable, indicating whether a person is likely to survive or die. The top 15 models each with an accuracy of over 70% were considered. The model will give important survival characteristics of a patient which include his history with diabetes, smoking, hypertension, and angioplasty.

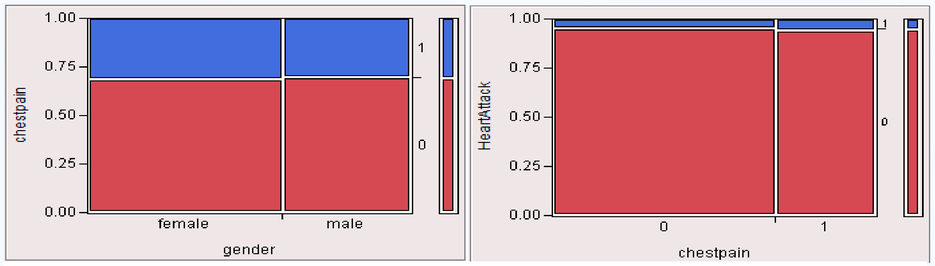
What is Myocardial infarction?

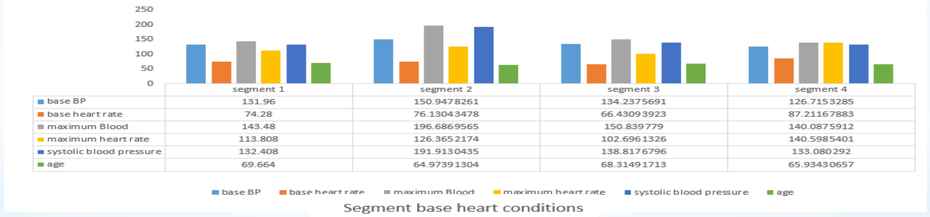
There are more people who die of heart disease in the United States than any other medical condition. This event occurs when blood stops flowing properly to a part of the heart which, in turn, affects the heart muscle as it does not receive enough oxygen. One of the coronary arteries that supplies blood to the heart develops a blockage due to an unstable buildup of white blood cells, cholesterol and fat.

Data Exploration

The dataset contains 1,380 records. Each record contains measurements for heart's performance such as maximum blood pressure, peak heart rate, base blood pressure, etc. with and without induced stress. This data is recorded for 544 men and 836 women. The survey to obtain the data was targeted for the older age group patients. Average age of the patients is 67.

Using the power of SAS® Enterprise Miner™ and its clustering algorithm, the patients are first segmented into different clusters each with its own stress characteristics. This is done in order to identify if there are any similar characteristics among a specific set of patients. These include base blood pressure, heart rate, and systolic blood pressure. The average basal heart rate for men and women are 74 and 76.13, respectively, which is higher as compared to the normal heart rate of 72



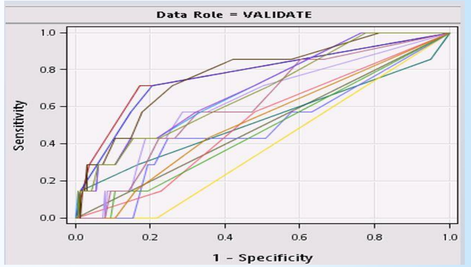


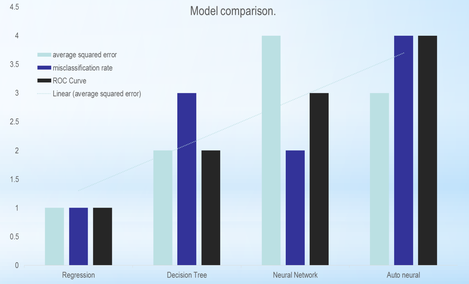
The factors that are highly responsible for cardiac arrest are the basal blood pressure, stress, and peak blood pressure. Once a patient is induced to stress, there is a rise in blood pressure which, in turn, leads to cardiac arrest. The variables basal BP, stress, and peak BP are highly correlated with chest pain. They represent a linear relationship as either one of them increase, the probability for a cardiac event to occur would also increase. Among all the variables, stress is the most important variable. When the patients are put under stress, 68% of the females and 69% of males experience chest pain. 95% of the patients who experience chest pain are likely to undergo cardiac arrest as seen from boxplot below. Irrespective of the gender, 95% of the patients experience cardiac arrest after being induced to stress. Males tend to experience a cardiac event at an average age of 64 and females at 69.

Modeling

Events that lead to a myocardial infarction have been analyzed. The survival of a patient after a cardiac event is predicted by a data mining model. The target variable is modeled as a binary variable with survival from the cardiac event and death as the two levels. The data set is split into 70% percent for training and the other 30% is used to validate the model.

The data mining algorithms of SAS® Enterprise Miner™ were used to build the predictive models. The regression models were built with various link functions in order to predict the probability that the target variable will obtain the event of interest as the function of one or more independent input variables. The decision trees were used to generalize the trends in data and are pruned in order to avoid over-fitting. As the Neural Network is a set of connected input/output units with each connection having a associated weight, it learns by adjusting the weights so as to be able to predict the correct class label of the input samples. The gradient boosting model was run to find the optimal split function for the best decision tree





Conclusions

The regression model predicts the outcome of a person's survival after a heart attack at a misclassification rate of 0.36 percent. 13% of the total patients survive after a heart attack. The patients who survive also did not undergo a recent bypass surgery, or an angioplasty surgery. The average blood pressure for patients who survive is 148.91, and the peak heart rate is 121. Over a 100 models were built, and the top 15 were selected and compared for their performances. 30% of the patients who survived did not experience a cardiac event before. The average age for survival is at 71, which is higher than the average age for whole data at 67. The power of SAS® Enterprise Miner™ was used to build a model, which yielded the most important characteristics for survival after a heart attack

References

http://support.sas.com/publishing/pubcat/chaps/57587.pdf

http://www.ats.ucla.edu/stat/sas/webbooks/reg/

Washington, D.C.

March 23–26, 2014