Sensyne Health

Machine Learning – Programming Assignment

April 2020

Sensyne Health

Designed by Clinicians. Focused on patients. Powered by AI.

1 Table of Contents

1 Introduction 2

2 Dataset 2

3 Assignment 2

4 Submission 3

1 Introduction

This assignment explores an interesting dataset and is purposely left vague. There isn't one single answer and it is up to you on how you would like to approach the problem. You have 48 hours to complete the assignment, but it's intended to take far shorter than that time.

Various options are given below so please pick one.

2 Dataset

Download the *processed.cleveland.data* file, which is available from the *Data Folder* at this link: https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

3 Assignment

Using Python or R perform one of the following tasks (using libraries that are generally accessible):

- Option 1: Predictive model. Use a predictive model to classify patients who have heart failure. Evaluate the performance of your model and suggest which features may be useful in heart failure prediction.
- Option 2: Statistical learning. Use statistical learning (with or without dimensionality reduction) to identify the combination of feature that are more likely to be associated with heart failure. Based on your analysis, suggest which features may be useful to predict heart failure.
- Option 3: Causal inference. Use causal inference to identify the features that are more likely to have a causal connection with heart failure. Based on your analysis, suggest which features may be useful to predict heart failure

- Option 4: If you would prefer not to apply the suggested approaches, then build another form of predictive/statistical model that will identify relevant features. Surprise us!

We'd like a brief report (no more than 2 pages including plots) outlining and justifying your choice of pre-processing, steps you may have taken to avoid overfitting, parameter choices and your choice of classifier / statistical modelling compared to other approaches. What can you conclude from this analysis? If your model is time consuming to train, then please provide the model as well.

Do reference any online code that you may have based your analysis on. We want to see your implementation of the analysis.

4 Submission

Please send us the code and report as follows:

1. Your 2-page report (as doc/pdf/txt)
2. Your code, either as html/pdf (for notebooks) or share their code in a pastebin platform such as https://hastebin.com.
3. Please DO NOT email us a compressed file (e.g. .zip, tar.gz, etc), as these often get blocked by our antivirus.

Please make sure you document the environment in which we should be able to run your code. This is to ensure that we can reproduce your results with minimum effort.

Finally, please keep everything confidential. Do not mention Sensyne Health or any obvious abbreviations in your public repositories / GitHub etc.