St. Michael's

Inspired Care.
Inspiring Science.

Li Ka Shing Centre for Healthcare Analytics Research and Training

Senior Data Scientist (Machine Learning) Take-Home Assignment

Time Allotted: 4 hours

Dataset and Tasks

You will be working with the PhysioNet 2012 Challenge dataset https://physionet.org/challenge/2012/ . The task and the dataset are similar to datasets you will be working with at LKS-CHART. You should work with the software package that you are most familiar with. You are free to use any code from the PhysioNet website. You are also free to use any code downloaded from the internet as long as it is not code written specifically for the PhysioNet 2012 Challenge dataset, and as long as the code is acknowledged. Please submit your code and your report, and save them in the following format: '*your-name_TestResponses*'.We should be able to reproduce all the results you report by running your code. Tests that are submitted after the 4 hour time limit will not be considered.

Tasks

1. Download and read in the dataset, and create a short report (1-2 pages, including figures, is sufficient) describing the dataset.
2. Create and evaluate a baseline algorithm for predicting death in the ICU. State precisely how you evaluated the algorithm.
3. Improve the algorithm by taking the fact that the patient's state changes over time into account. Evaluate your improvement.
4. Devise a method to determine which variables are particularly useful for predicting mortality in the ICU. Report your results.
5. Devise and try out a method that finds complex features that are useful for predicting mortality in the ICU. Report your results.
6. Formulate a plan for working with this dataset further. Include this plan in your report.