pacmed

Application for Data Scientist

Assignment

Who can be discharged from the Intensive Care Unit?

Application for Data Scientist: Assignment

Assignment

This document contains the assignment that is part of the Data Scientist application process.

You are free to use as much time as you want, but we do not expect you to invest more time than one working day.

To do the assignments, we prefer you to use Python and specifically Jupyter Notebooks. Make sure that your code is reproducible and easy to follow.

You can send your solution to daan.debruin@pacmed.nl.

Introduction: Discharging patients from the Intensive Care Unit

Background

At the Intensive Care Unit (ICU), patients who are in critical, life-threatening conditions receive continuous care. The purpose of the ICU recording is to bring the patient back to a stable condition, after which they can go back to the normal ward. The moment of discharge is important: discharging a patient too early can lead to complications and being *readmitted* to the ICU with increased risk of death. Keeping patients too long is also undesirable, capacity is limited and patients can get complications as a result of being at the ICU for too long. Can Machine Learning help IC doctors decide when a patient can be discharged?

Data

Enormous amounts of data are continuously collected at the IC. Physiological values such as heart rate and blood pressure are recorded per minute, and in addition, patient characteristics, clinical observations and laboratory results such as blood values are known. You will receive three (artificial) datasets with IC data.

age.csv

This is a file containing the patients age, it contains the fields

- *pat_id:* the id of the patient
- *age*: the age of the patient today

admission.csv

This is a file containing information on when a patient is admitted to and discharged from the IC. A patient can enter multiple times in this dataset.

- *pat_id*: the id of the patient
- *date_admission*: the date the patient is admitted to the IC
- *date_discharge*: the date the patient is discharged from the IC

signal.csv

this is a file containing (artificial) information on three physiological parameters of the patient.

- *pat_id*: the id of the patiënt
- *day*: the day the signal is measured
- *hour*: the hour the signal is measured
- *parameter*: the parameter that has been measured
- *value*: the value of the measurement

Assignment: who can (not) be discharged?

Assignment

- Build an algorithm that can help the doctors at the IC decide who can be discharged using the signal data of the patient. This algorithm should be able to predict which patients have a high risk of being readmitted if they where to be discharged.
- *Note that in real life, there are hundreds of different parameters that have been measured. In the case you will receive three of these parameters. Therefore your model will probably not have a very high accuracy, but you should be able to identify patients that are at risk.*
- Your output should include reproducible code that transforms the raw data into the final algorithm, including all the relevant steps you took in making this algorithm

Scoring

Your solution will be assessed on the following points. We will not expect you to excel on all of these points, but you are given the freedom to show off in these directions:

- Code quality
- Descriptive analyses
- Feature and outcome (X and y) engineering
- Professional machine learning pipeline
- Model performance and interpretation