In order to solve this task we first loaded the data into the code and then copied the training labels 12 times so that the train\_features and train\_labels have the same amount of rows. We did this in order to use all 12 hours of data of every patient. We then declared the needed labels for each subtask. For each subtask we dropped the pid and the time column as we didn’t think that they would be relevant. Furthermore, we replaced every NaN with the mean of each column, as we thought that would make most sense. Replacing a NaN with zero wouldn’t make sense in most contexts (e.g. heartrate or oxygen concentration).

For subtask 1 we used a random forest classifier. We experimented a lot with different classifiers and even ANNs but found that we got the best results with the mentioned classifier. We set the amount of estimators arbitrarily to 300 and didn’t set class\_weight, even though the data is imbalanced. We found that this gives us better results. We fitted the data and then printed the ROC-AUC score for each column. Lastly, we predicted the probabilities of test\_features and took the mean of the 12 values we got for every hour.

In subtask 2 we did the same thing as for subtask 1.

In subtask 3 we experimented with previously learned regression techniques and experimented with Lasso and Ridge with cross-validation. We found that Ridge gave us the best results. We used the MSE as a metric as we thought that would be a good one for a regression. We then predicted the needed labels again, using the mean of the 12 hours.

Finally, we wrote everything into one dataframe and submitted our solution.