techolution

Assignment for Data Science Interns

The dataset shared herewith is from the informatics department of a hospital which has patient level vitals for each patient they have on ICU beds and the corresponding alerts based on the status of their experiencing a code-blue event (viz. "Coded" = "high-risk":1 or 'low-risk' :0). The Schema for this dataset is shared below :

Features and Description:

vitals_datetime
Time at which the patient got admitted to the hospital

heart_rate
Patient vital

respiration_over_impedence
Patient vital

spirometry_oxygen_saturation
Patient vital

pulse
Patient vital

blood_pressure_systolic
Patient vital

blood_pressure_diastolic
Patient vital

blood_pressure_average
Patient vital

patient_id
Unique-id of the patient

machine_id
Unique machine id (ICU bed ID)

Coded
Actual alerts [high-risk/low-risk]

techolution

The dataset is divided into two csv files training_frame.csv for training a model that predicts the status of a code-blue event either on the day-of or one day ahead of the event. The test_frame.csv is meant for testing out-of-sample.

- Build models to predict the Coded column of the training frame and test the model on out of sample data i.e. test_frame, after aggregating the data by patient, by day.
- Mark a patient-day combination as a high-risk event if there was at least one code-blue event on that day.
- Develop multiple models and choose the best one or build an ensemble of these models to determine code-blue risk.
- Document your code and approach.

Submit your code as a Jupyter notebook that can run on Google Colab. Submit “assignment.ipynb” file to us.