Predicting heart rate to monitor stress level

Business Context

Anxiety and stress make your heart work harder. When you're under stress your body's "fight or flight" response is triggered i.e. your body tenses, your blood pressure rises and your heart beats faster. Stress hormones may damage the lining of the arteries. In the current scenario post-covid, since most of us are indoors, stress levels are at an all time high due to increasing anxieties which is leading to a higher heart rate. And your body's response to stress may be a headache, back strain, or stomach pains. Stress can also zap your energy, wreak havoc on your sleep and make you feel cranky, forgetful and out of control.

Higher heart rate is not always better since pathological conditions can lead to an increased heart rate. Tachycardia refers to a fast resting heart rate, usually over 100 beats per minute. Tachycardia can be dangerous, depending on its underlying cause and on how hard the heart has to work.

An optimal level of heart rate is associated with health and self-regulatory capacity, and adaptability or resilience. Higher levels of resting vagally-mediated heart rate are linked to performance of executive functions like attention and emotional processing by the prefrontal cortex.

Higher heart rates are usually connected with higher stress levels. When stress is excessive, it can contribute to everything from high blood pressure, also called hypertension, to asthma to ulcers to irritable bowel syndrome

Stress may affect behaviors and factors that increase heart disease risk: high blood pressure and cholesterol levels, smoking, physical inactivity and overeating. Some people may choose to drink too much alcohol or smoke cigarettes to "manage" their chronic stress, however these habits can increase blood pressure and may damage artery walls.

Thus, heart rate can be used to monitor your stress levels and keep it under check as it is a useful indicator of good health.

A recent study speaks about effects of stress on increased heart attacks amongst 30-40 year olds:

https://economictimes.indiatimes.com/magazines/panache/heart-attacks-on-the-rise-among-30-40-year-olds-diabetes-hypertension-are-contributing-factors/articleshow/66997025.cms

About the Data

The data comprises various attributes taken from signals measured using ECG recorded for different individuals having different heart rates at the time the measurement was taken. These various features contribute to the heart rate at the given instant of time for the individual.

You have been provided with a total of 7 CSV files with the names as follows:

time_domain_features_train.csv - This file contains all time domain features of heart rate for training data

frequency_domain_features_train.csv - This file contains all frequency domain features of heart rate for training data

heart_rate_non_linear_features_train.csv - This file contains all non linear features of heart rate for training data

time_domain_features_test.csv - This file contains all time domain features of heart rate for testing data

frequency_domain_features_test.csv - This file contains all frequency domain features of heart rate for testing data

heart_rate_non_linear_features_test.csv - This file contains all non linear features of heart rate for testing data

sample_submission.csv - This file contains the format in which you need to make submissions to the portal

Following is the data dictionary for the features you will come across in the files mentioned:

MEAN_RR - Mean of RR intervals

MEDIAN_RR - Median of RR intervals

SDRR - Standard deviation of RR intervals

RMSSD - Root mean square of successive RR interval differences

SDSD - Standard deviation of successive RR interval differences

SDRR_RMSSD - Ratio of SDRR / RMSSD

pNN25 - Percentage of successive RR intervals that differ by more than 25 ms

pNN50 - Percentage of successive RR intervals that differ by more than 50 ms

KURT - Kurtosis of distribution of successive RR intervals

SKEW - Skew of distribution of successive RR intervals

MEAN_REL_RR - Mean of relative RR intervals

MEDIAN_REL_RR - Median of relative RR intervals

SDRR_REL_RR - Standard deviation of relative RR intervals

RMSSD_REL_RR - Root mean square of successive relative RR interval differences

SDSD_REL_RR - Standard deviation of successive relative RR interval differences

SDRR_RMSSD_REL_RR - Ratio of SDRR/RMSSD for relative RR interval differences

KURT_REL_RR - Kurtosis of distribution of relative RR intervals

SKEW_REL_RR - Skewness of distribution of relative RR intervals

uuid - Unique ID for each patient

VLF - Absolute power of the very low frequency band (0.0033 - 0.04 Hz)

VLF_PCT - Principal component transform of VLF

LF - Absolute power of the low frequency band (0.04 - 0.15 Hz)

LF_PCT - Principal component transform of LF

LF_NU - Absolute power of the low frequency band in normal units

HF - Absolute power of the high frequency band (0.15 - 0.4 Hz)

HF_PCT - Principal component transform of HF

HF_NU - Absolute power of the highest frequency band in normal units

TP - Total power of RR intervals

LF_HF - Ratio of LF to HF

HF_LF - Ratio of HF to LF

SD1 - Poincaré plot standard deviation perpendicular to the line of identity

SD2 - Poincaré plot standard deviation along the line of identity

Sampen - sample entropy which measures the regularity and complexity of a time series

higuci - higuci fractal dimension of heartrate

datasetId - ID of the whole dataset

condition - condition of the patient at the time the data was recorded

HR - Heart rate of the patient at the time of data recorded

Objective

The objective is to build a regressor model which can predict the heart rate of an individual. This prediction can help to monitor stress levels of the individual.

Evaluation Metric

Mean Absolute Error :

$$\mathbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{x}_i - \boldsymbol{x}|$$

n - total number of predicted samples, x_i predicted output, x actual output

Submission Process

- You are required to submit a csv file which contains the uuid and its predicted label(HR).
- Please note that file should be in a csv format as shown in sample_submission.csv
- Please ensure that submission file contains all the test instances