General Championship Data Analytics

Problem statement

An automated manufacturing unit uses 25 sensors to record data and these data are then used to predict the probable state of the unit. The sensors record the data in every minute and in every 10 minutes the state of the unit is also recorded. The state could be ‘no risk’, ‘low risk’, ‘medium risk’ and ‘catastrophic’. The goal is to build model that will take the data from 25 sensors and predict the state of the unit in every 10 minutes.

Training data format: The traning data is given in a file called “train.txt”. The file is comma separated. The first and the second field in each line give the sample id and the timepoint respectively. The other fields in the line are the normalized values from the sensors. After every 10 line, you have a line with only one integer, which is the state label of the unit. The integer value 0, 1, 2, 3 refer to no risk’, ‘low risk’, ‘medium risk’ and ‘catastrophic’ respectively.

Test data: The file named ‘test.txt’ contains the test data. Once again, the data has the same format as the train data but without the true labels. The participants need to submit a .txt file with two columns (comma separated), where the first column is the sample id and the second column is the label (the state of the unit).