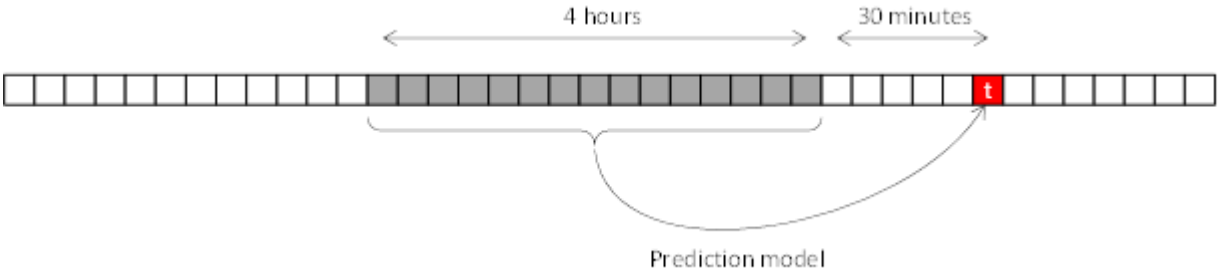
From: Gutierrez-Osuna, Ricardo <rgutier@cse.tamu.edu>
Sent: Monday, May 22, 2023 4:53 PM
To: Panwar, Sharaj <sharaj@exchange.tamu.edu>
Subject: RE: Request to consider in your Research Group

Sounds good.

Attached is the dataset. You will find 113 files. Each file contains measurements from a separate sensor, taken every 5 minutes for about a month. Your goal is to predict the sensor measurement at time [t] from a 4-hour window of past measurements starting 30 minutes earlier: [t-30, t-35, t-40,…], as illustrated below:



I want you to validate the models using an 80:20 split, where you randomly select 80% of the sensors (i.e., files) for training and the remaining 20% of the sensors for testing. In other words, none of the data from the test sensors can be used for training, the models are tested across sensors. The measure of performance is root mean squared error (RMSE) of the prediction on the test sensors. You may want to repeat the 80:20 split multiple times, and report the average.

I will be evaluating your solution both qualitatively (how well though-out your approach is) and quantitively (RMSE). You can use any machine learning model and any libraries that you wish. The deadline is this Wednesday at 4pm.

Let's see what you can do!

Ricardo