

For odd Roll Numbers:

Use the file Shared.xlsx. The file has three sheets: Training, Validation and Test. Data in these sheets are stream data. Each row represents values measured by 4 sensors. Every day only one measurement is taken by the set of sensors. Data in the Training sheet is to be used to train the model while data in the Validation sheet is provided to validate the model. In no case, data in the Validation sheet should be used for training the model. Use Approximate Stream Regression to develop a regression model which can predict the value of variable D. I1, I2 and I3 are independent variables (you are free to derive more independent variables using I1, I2 and I3. It is also not mandatory to use all three independent variables). Window size can be any value between 10 to 100 examples. You can tune the window size and other hyper parameters of the algorithm for the given task and data. Using this developed model make prediction for each observation (there are 108 observation) in the test set. Store this prediction in a .csv file (This file would have only one column without header and 108 predictions/rows in this column). Name of these .csv file should be your roll number in BLOCK letters. Submit this file. I have held back values of dependent variable for the observations in the test set. I will compare them with your predictions to compute Mean Absolute Error. Also, Submit .ipynb file (Name of these .ipynb file should be your roll number in BLOCK letters) with proper commenting where ever required. If I wish, I should be able to run the .ipynb file and make predictions for the observations in the Test sheet on my own (which will stored in YOURROLLNO.csv) assuming that Shared.xlsx and .ipynb file are in the same directory. Please note, if I will find a batch implementation, 0 marks would be awarded.

Deadline for Submission: 12:45 pm. I will share a mechanism (most probably, a Google form) for the submission before the deadline.