From Vidushi Shukla to Everyone: 11:08 AM

Can you discuss the ROC-AUC curve?

From Tianqi Tang to Everyone: 11:13 AM

Are other fractions ok? For example, I used, 0.64, 0.16, 0.2

From Hardy Bright to Everyone: 11:18 AM

Would you talk a little more about concatenating the dataset? I missed the fact that using the different preprocessors actually splits the set.

From Vidushi Shukla to Everyone: 11:21 AM

Is it ok to use simply stratify to split instead of StratifiedKFold?

How would we stratify using both User and Class? I thought it only allows for one column?

From Connor Uzzo to Everyone: 11:22 AM

How would you split non IID data for a regression problem? would the same strategy work?

From Hardy Bright to Everyone: 11:23 AM

I know that we will likely not have a solution set. Would the TAs be able to work with us to debug coding issues?

From Tianqi Tang to Everyone: 11:24 AM

When I used one hot encoder it creates more columns. And the encoders convert from dataframe to array. Is this expected?

From Vidushi Shukla to Everyone: 11:29 AM

Can we ask questions about PS5?

From Connor Uzzo to Everyone: 11:30 AM

should we not do any sort of stratification if we have time series data then?

From Vidushi Shukla to Everyone: 11:31 AM

Problem 2 is for 16 points which made me think we have to write a long explanation (4 points for each subpart). But for the two cases with dropping columns and rows with missing values, the explanation is basically just "Don't do this, too much data is lost." Is there any other reason not to do this?

From Christopher Rohlicek to Everyone: 11:32 AM

In terms of generating features for our report, should we standardize our features before creating our figures plotting features against each other?

generating figures\*\*

Do not preprocessing the target variable!!!!

