Scenario: A research team reports that they can reliably predict whether a participant is politically conservative or liberal by how their brain responds to images on a computer screen. Participants are shown a series of photos of 30 different inanimate objects (e.g., cars, mountains, rocks) while in an MRI system and asked to think about their meaning. The hemodynamic response to each image is measured in a set of 20,000 voxels. The authors use a leave one out cross validation approach, by training a naive Bayes classifier (which is relatively insensitive to $p>n$ dimensions in $\mathbf{X}$) on brain responses of all $n-1$ participants and predicting the hold out test participant's political affiliation.

The authors select the features for their analysis in two ways. They first measure the stability of voxel responses (by measuring the trial-by-trial variability) and only choose to analyze the 200 voxels that are deemed most stable (i.e., have the most consistent responses across trials). Voxel stability data for the hold out test subject is not included in the stability analysis on each fold of the LDA classifier. Using a forward stepwise selection procedure on whole brain voxelwise responses, the authors also reduce the number of images used by their classifier from 30 to 6. The average response from the same 6 images is used for all runs of the classifier.

Question: Did the authors commit feature hacking? If so, how?