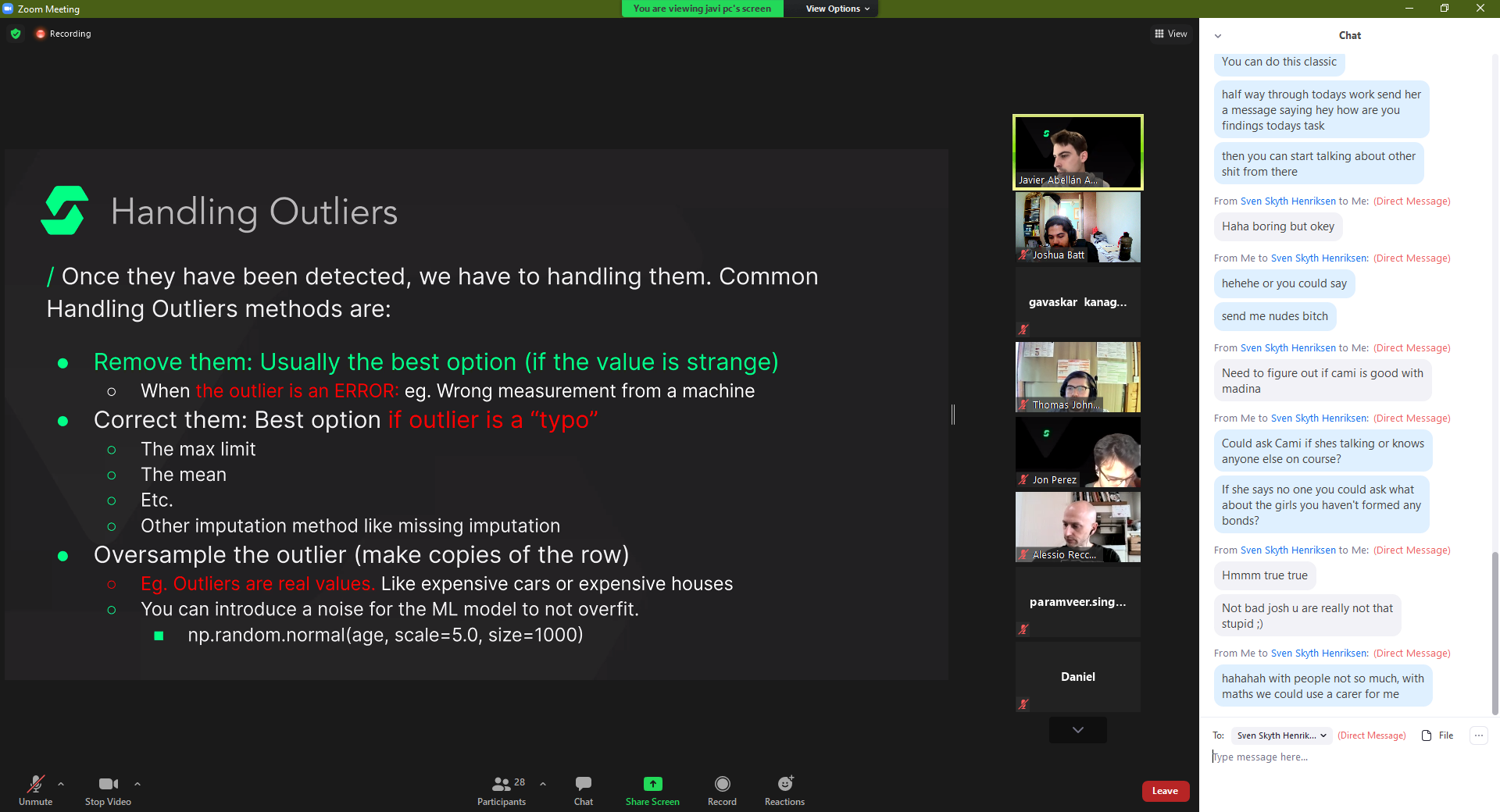
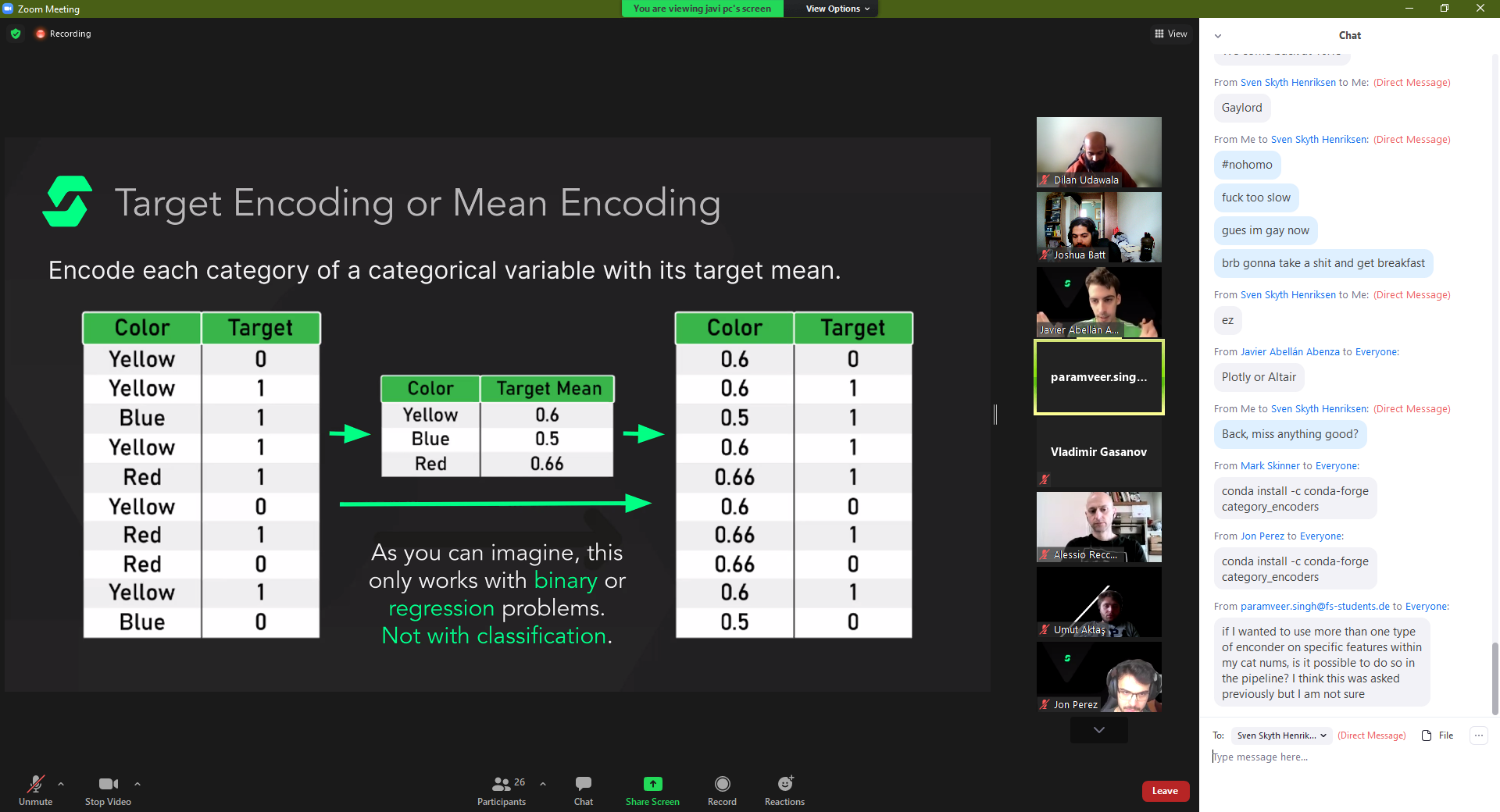
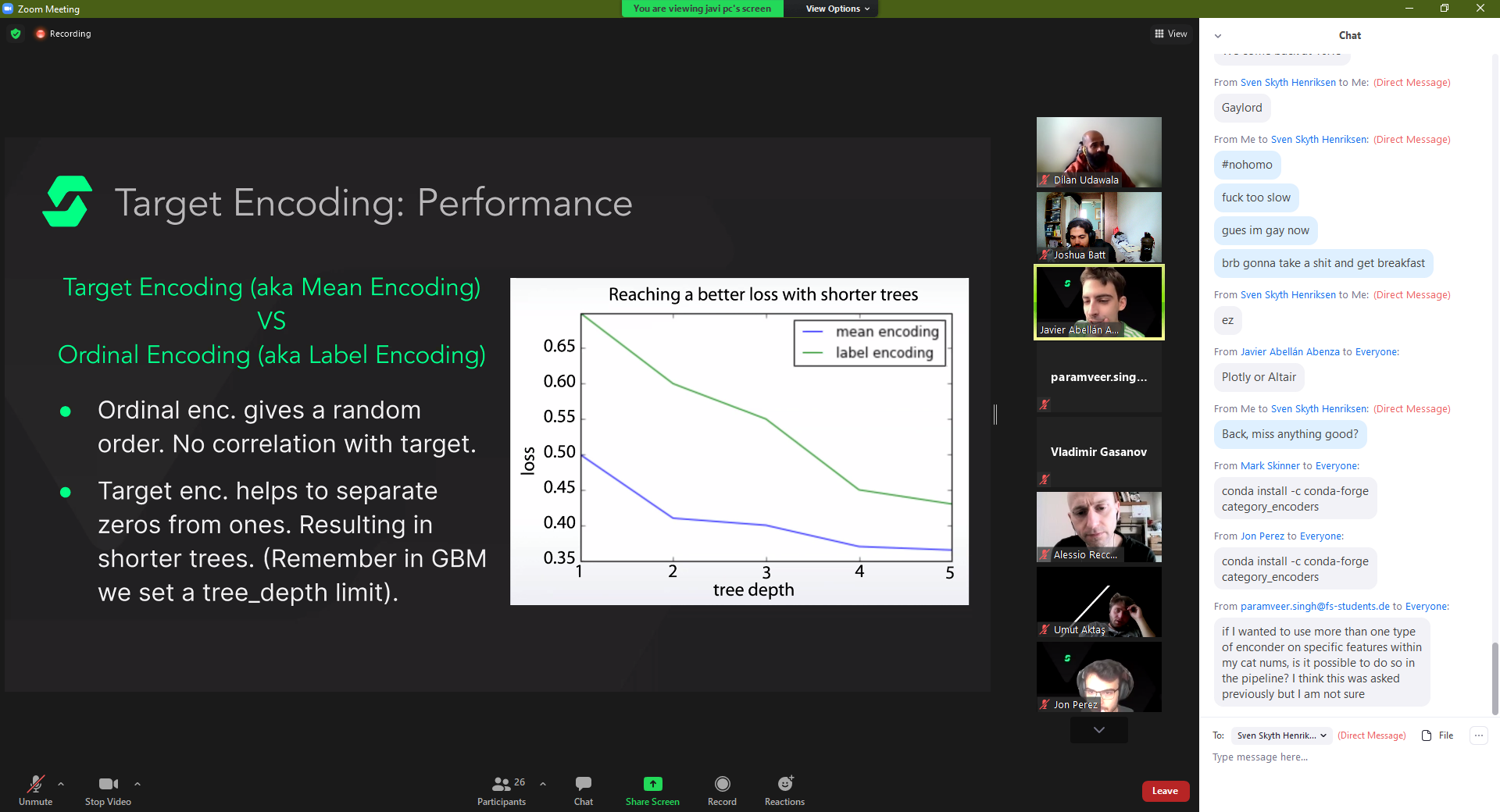
**M4 D4**

**Outliers:**

* Data points that strongly deviates from the norm. Should be removed to not confuse the ML model.
* Manual method to detect outliers is to plot the data. Easiest graph to spot outliers is the scatter plot.
* Fuzzywuzzy library tests categorical data (words) similarity.
* Handling categorical data:
  + 
* Hyperparams is changing the parameters by a logarithmic scale (instead of going from 1 to 2. You do 1 to 1.1. If too much, then 1 to 1.01. etc…)
* Fake data is made to generate noise that follows the patten of the data. This will help the model not to over fit the data.

**Categorical Encoders:**

* Most common type
  + 
  + How this mean value encoding works visually:
    - 
* Target Encoding:
  + 
  + k