

Hurdles:

1. Computational Constraints: I am only using 2 lakhs observations out of 5.3 million. Do you think it is a wise decision considering computational cost or shall I try moving towards some cloud computing options?
2. I am trying to extract vegetation data like NDVI. Last time I tried, I got 6,000 rows only and thus that is a big limitation. So, for my upcoming days, shall I focus on exploring more dataset options or enhancing my model?

Questions:

1. Since my target variable is very imbalanced, would you recommend me to explore more techniques like SMOTE, or my approach stratified sampling is fine?
2. Do we have to submit our codes for the final assessment?
3. For the draft report, how much would you expect us to complete? Also, should it reflect my current results ?