6th place - 1 week rush

Wow! I am impressed that I somehow managed to somewhat make a comeback.

My Scuffed Solution

Early on I think everyone noticed CatBoost was really bad, so I never used CatBoost in my ensemble. My ensemble was only made of XGBoost (70% of the oofs) and LightGBM (30% of the oofs). I collected 135 oofs and preds that were all created by myself, because one thing I learned was that a single bad oof or pred prediction sourced from a public notebook without careful check could result in a really bad leaderboard score. My final model architecture was a simple mean between a logistic regression and hill climb of all my oofs. I trained all my models locally on my 4070 super, works really well.

Some Secrets I might have gatekeeped

I discovered that setting all the data types to categorical improved the score early on
Adding original data also drastically improved the score
XGBoost sampling_method = gradients somehow worked better even with high subsample?
XGBoost refresh_leaf = 0 also slightly improved the score, might be overfitting

Things that sucked

CatBoost
Feature Engineering
Slow Training Times, I think 95% of the models use a learning rate of 0.05 😅
Summer School starting so I only had like 1 week before I couldn't prioritize coding anymore
Logistic Regression taking 3 hours to run on the CPU
Adding additional OOF predictions and somehow the Hill Climbing score becomes worse

Happy Summer Break

Unless you also have summer school

PS

How do you know if you won a prize?