So here is the data set the response are the mean losses 2 and the predictors are width and Undercladding material ,Waveguide material and Uppercladding material . so here i need a train test split of the first 4 wafers in every material combination as train data and the 5 test. so wafer id 1-4 train 5 test 6-9 train 10 test and so on. then do in R a mlr with all interactions and remove gardually start wit a rather complex predictor (predictor1+predictor2 + predictor3+predictor4)^2 and to stepwise eliminate single
terms without violating a hierarchical structure (drop1 procedure) and then test with the testset and use MSE, RMSE and MAE
then train here a CART model with bayesian hyperparameter optimization and 10 fold crossvalidation . Then use MSE, RMSE, MAE on the predicted vs true values.