

Our team's goals for the next few weeks:

A. Optimizing

1. Get EBM optimized
 - a. Increase max depth until accuracy starts to increase
 - b. Increase number of base models until max accuracy
2. Get LGBM optimized
 - a. Increase max depth until accuracy starts to increase
 - b. Increase number of base models until max accuracy
3. Compare performance of optimized models
 - a. MSE and post hoc analysis (PDP, ICE, etc)

B. Overfitting

4. Overfit EBM
 - a. Max depth and number of base models very high
5. Overfit LGBM
 - a. Max depth and number of base models very high
6. Compare overfitted models
 - a. MSE and post hoc analysis (PDP, ICE, etc)

C. Underfitting

7. Underfit EBM
 - a. Max depth and number of base models very low
8. Underfit LGBM
 - a. Max depth and number of base models very high
9. Compare underfitted models
 - a. MSE and PDP

D. Stretch Goals

10. Model Robustness
11. Adding in λ_1 λ_2

Results

Models	Train MSE	Test MSE	Gap MSE
Optimized EBM			
Optimized LightGBM			
Overfit EBM			
Overfit LGBM			
Underfit EBM			
Underfit LGBM			

Models	Max depth (max leaves)	Number of base models (interaction)
Optimized EBM		
Optimized LightGBM		
Overfit EBM		
Overfit LGBM		
Underfit EBM		
Underfit LGBM		

Our final report will generally follow this outline:

1. Introduction
 - A. Client
 - B. Problem Statement
 - C. Intent of Report & Relevance of Comparing the Models (What models did we choose? Why did we choose these models? What do they tell us, making them the most significant to compare? What is our goal (bridge back into how we are addressing the problem statement?) (Our goal is to determine if the switch to black box model is necessary or if interpretable models suffice due to difficulties in using black box models in the financial industry).
2. Background Information
 - A. What are Interpretable Models?
 - B. Pros & Cons of Interpretable Models
 - C. What are Black Box Models?
 - D. Pros & Cons of Black Box Models
 - E. Interpretable Models versus Black Box Models
 - F. Plots that explain the results of each of the models
 - G. Explain each of the plots and what they show (significance of their results to model robustness to showcase why they are selected to begin with)
 - H. Pros & Cons of each plot
 - I. Bridge into the models we have chosen to work with and why
3. Light GBM
 - A. Introduction: what is light GBM and why is it important?
 - B. How things are calculated within Light GBM - what parameters are considered
 - C. Bridge into why optimization is necessary using some information about problems if not optimized (overfitting and underfitting)
 - D. Process of optimization
 - E. Show results proving why optimization adds to model robustness and efficiency (comparison between optimized, overfit, underfit)
 - F. Show various plots explaining robustness and efficiency
4. EBM
 - A. Introduction: what EBM is and why is it important?
 - B. How things are calculated within EBM - what parameters are considered
 - C. Bridge into why optimization is necessary using some information about problems if not optimized (overfitting and underfitting)
 - D. Process of optimization
 - E. Show results proving why optimization adds to model robustness and efficiency (comparison between optimized, overfit, underfit)
 - F. Show various plots explaining robustness and efficiency
5. Model Comparison
 - A. Compare the robustness and efficiency of each model relative to one another using accuracy measures like MSE