Solution Summary

To reproduce score:

1. Upload attached notebook to colab,
2. Upload test data
3. Run all to get submission file

High level summary of techniques used

Outlier Treatment:

My solution greatly benefited from eliminating outliers. I removed target values that were below 20 AGBD and above 300 AGBD.

Feature Engineering:

Cloud Masking:

Both the training and testing images exhibited a significant presence of clouds. To address this issue, I generated a cloud mask and replaced the corresponding pixels in the training and testing images with their mean values.

Vegetation Indices:

I generated a total of 187 distinct vegetation indices. The top 10 vegetation indices, identified through model feature importance, are as follows:

1. NMDI - Normalized Multi-band Drought Index.
2. S2REP - Sentinel-2 Red-Edge Position.
3. SIPI - Structure Insensitive Pigment Index.
4. TTVI - Transformed Triangular Vegetation Index.
5. MTCI - MERIS Terrestrial Chlorophyll Index.
6. NHFD - Non-Homogeneous Feature Difference.
7. NBRplus - Normalized Burn Ratio Plus.
8. MIRBI - Mid-Infrared Burn Index.
9. IBI - Index-Based Built-Up Index.
10. AWEInsh - Automated Water Extraction Index with Shadows Elimination.

Training and validation sets:

The training data was divided into 25 folds, with the sixth fold utilized as the validation set, and the remaining 24 folds used for training.

Modelling:

The model employed for this solution was Lightgbm, with the objective set to quantile.