We have used regression technique to accomplish this task. Under this we leveraged 2 algorithm first one is linear regularized model (LASSO) and second one is non-linear (Lightgbm), We combined both for the final prediction. Whole procedure is easily understandable, please have a look at the below work flow diagram.

Data

Algorithm

Feature Engineering

Lasso

LightGBM

0.5 0.5

Feature engineering: Under this we have performed the following tasks.

* Attributes formatting (HS).
* Deals with invalid values (Avg).
* New feature generation.
* Feature selection ( correlation, univariate).

Model: In modelling, first we splitted dataset in training and validation set (25%). After **validating** both the model separately based on **Mean absolute error** (mea) we merged them with 50% weightage to each.

Pipeline: Finally we have built a pipeline for all the pre-processing task including final prediction.