**Model Explaination**

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 model for the final prediction, combining model minimizes the error. Whole procedure is easy understand, 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 (Like Runs scored through boundary)
* Feature selection (correlation, univariate).

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

**Pipeline**: Finally we have built a pipeline for all the pre-processing task including final prediction. As we do not need to repeat same process for the test dataset also, pipeline is very much helpful.

Thank You

Regards

Ranjeet Kumar

Pankaj Kumar

Sourav Suman