Energy prediction using Ensembling Techniques on the ASHRAE dataset

BY: Sahil Bansal
Juhi Checker

Conventional Approach

Technical challenges of building a single estimator:

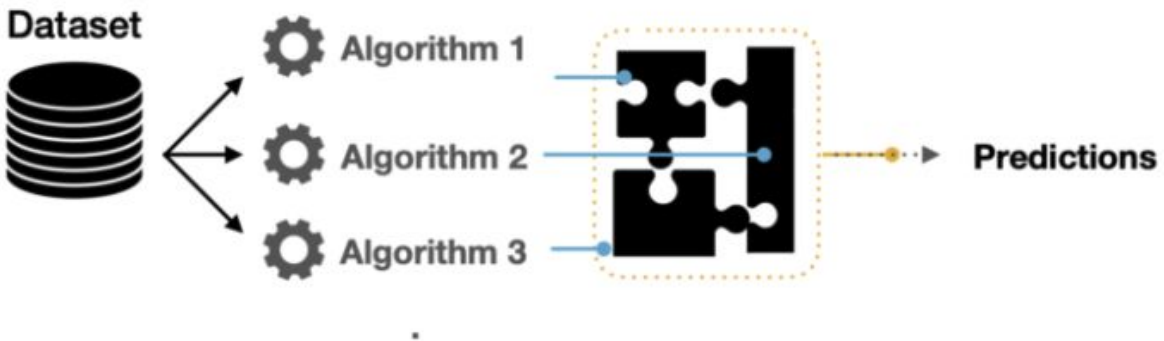
High variance: The model is very sensitive to the provided inputs to the learned features.

Low accuracy: One model or one algorithm to fit the entire training data might not be good enough to meet expectations.

Features noise and bias: The model relies heavily on one or a few features while making a prediction.

ENSEMBLING

Ensemble methods is a **machine learning technique** that combines several base models in order to produce one optimal predictive model.



https://towardsdatascience.com/ensemble-models-5a62d4f4cb0c

ENSEMBLING TECHNIQUES

- Bagging
- Boosting
- Blending
- **Stacking**
  - Mean
  - Median
  - TensorFlow Model

ASHRAE DATASET

The Ashrae data comes from over 1,000 buildings over a three-year timeframe. We are given 3 different Datasets which can be used for training. Our task is to develop accurate models to predict the energy usage of the building.

	building_id	meter	timestamp	meter_reading
200	173	1	2016-01-01 00:00:00	19.3427
201	174	0	2016-01-01 00:00:00	179.9000
202	174	1	2016-01-01 00:00:00	52.8583
203	175	0	2016-01-01 00:00:00	86.5900
204	175	1	2016-01-01 00:00:00	116.3370

train.csv

	site_id	building_id	primary_use	square_feet	year_built	floor_count
0	0	0	0	7432	2008.0	NaN
1	0	1	0	2720	2004.0	NaN
2	0	2	0	5376	1991.0	NaN
3	0	3	0	23685	2002.0	NaN
4	0	4	0	116607	1975.0	NaN

building_metadata.csv

Results

Root Mean Squared Logarithmic error

Half and Half- 1.0053

Light BGM- 0.9855

Highway Kernel- 0.9966

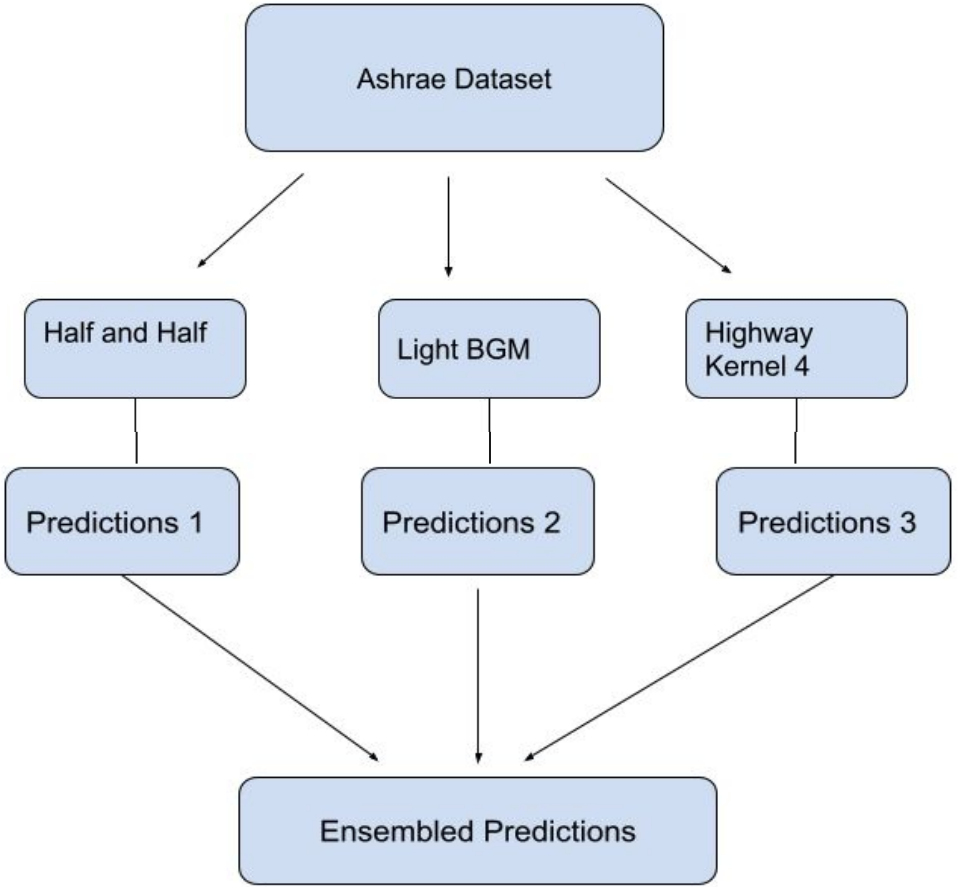


Figure 1: Flowchart for training

Results

$$Y = w1*pred1 + w2*pred2 + w3*pred3 \qquad eq(1)$$

Root Mean Squared Logarithmic error

Mean- 0.9776

Median- 0.9816

TensorFlow SGB Model- 0.9760

w1, w2 and w3 are weights found through TensorFlow Model.

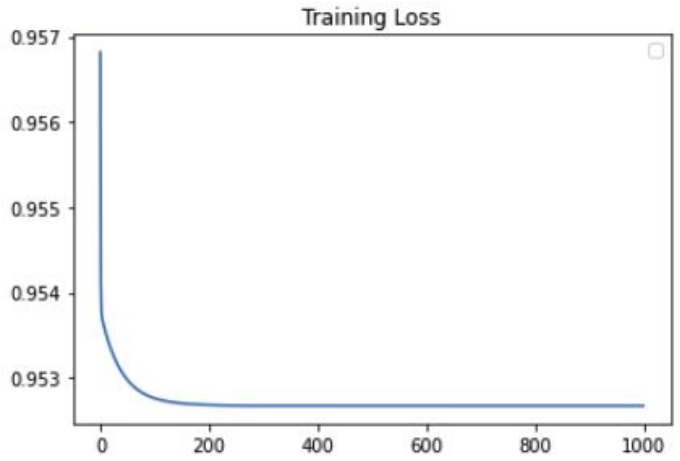


Figure 2: Training loss graph