kaggle DAYS MEETUP

# ASHRAE - Great Energy Predictor III

## Hyperparameter Optimization and Stacking

Siddharth Yadav

INTERNSHALA kaggle LOGICNI Devfolio

# Agenda

kaggle DAYS MEETUP

- About me
- Kaggle Kernels - Why you should use them?
- About the ASHRAE competition
- Current trends and tips for the competition
- Using Bayesian Optimization - Why? When? How?
- Stacking - The rewards of blending
- Q&A

INTERNSHALA kaggle LOGICNI Devfolio

# About me

kaggle DAYS MEETUP

- Elitist nerd due to circumstances
- B.Tech Final Year Student
- Kaggle Kernels Master
- Passionate for all things “Data”
- Got into Data Science just before the hype
- Worked for organisations like

MHRD Govt. of India
सत्यमेव जयते

Fidelity INTERNATIONAL

INTERNSHALA kaggle LOGICNI Devfolio

# Kaggle Kernels

Jupyter Notebooks that run on a hosted runtime on kaggle free of charge.

You don't have to worry about storage, processing or other resources.

You also get GPU and version control on the platform itself. importantly,

Most important component of any platform: good community support.

INTERNSHALA kaggle LOGICNI Devfolio

# Why use kernels?

Github works for developers. but how to showcase your skills in analytics and statistics.

Kernels provide more exposure and feedback.

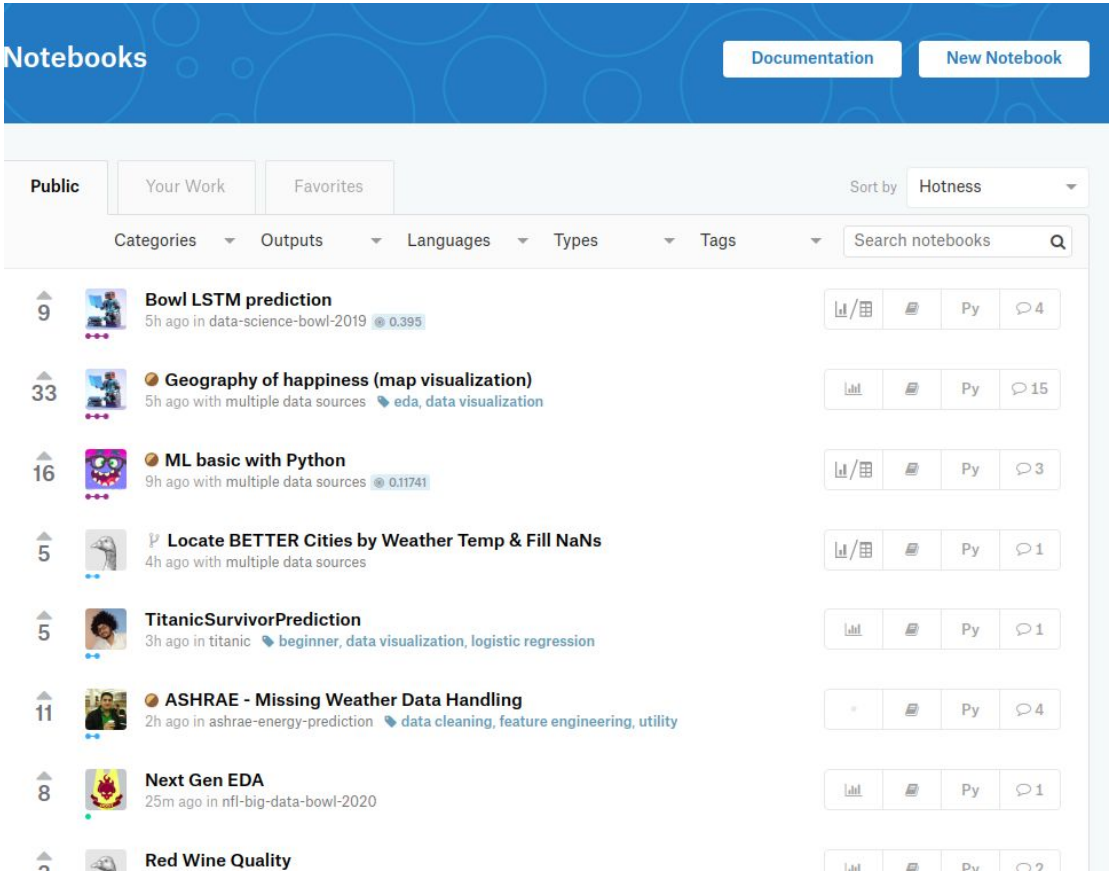
High end kernels make good projects for your resume. Helped me nail an interview. True story.

The stuff you learn through kernels won't be found in books or online courses.

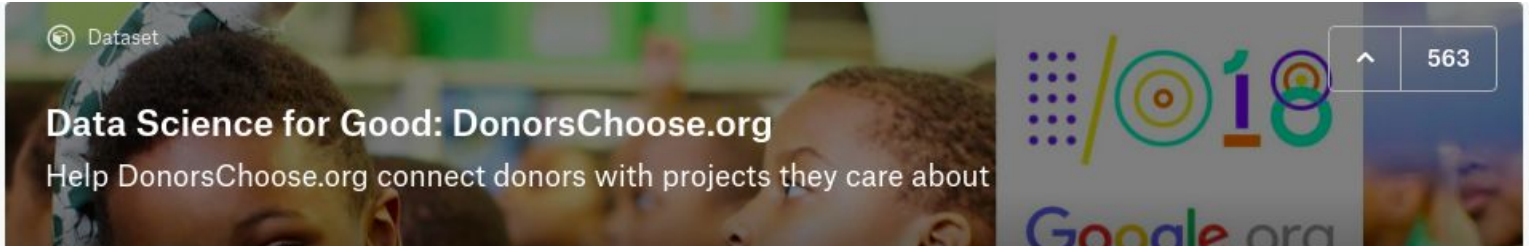
INTERNSHALA kaggle LOGICNI Devfolio

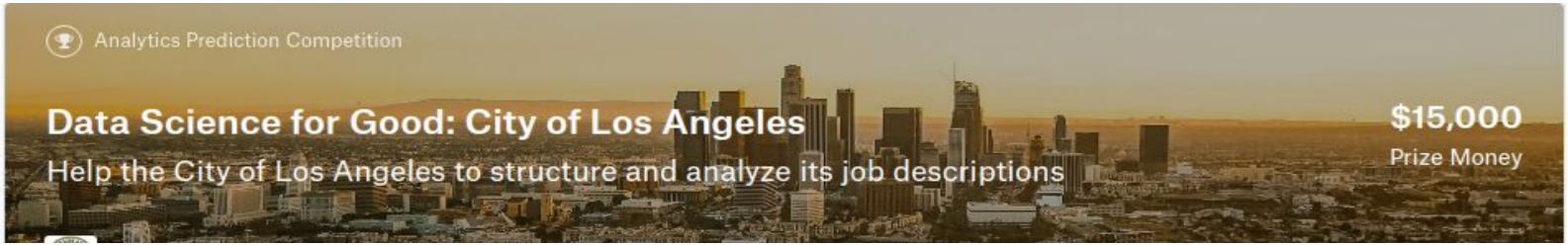
How to approach kernels?
Easy. Explore them.

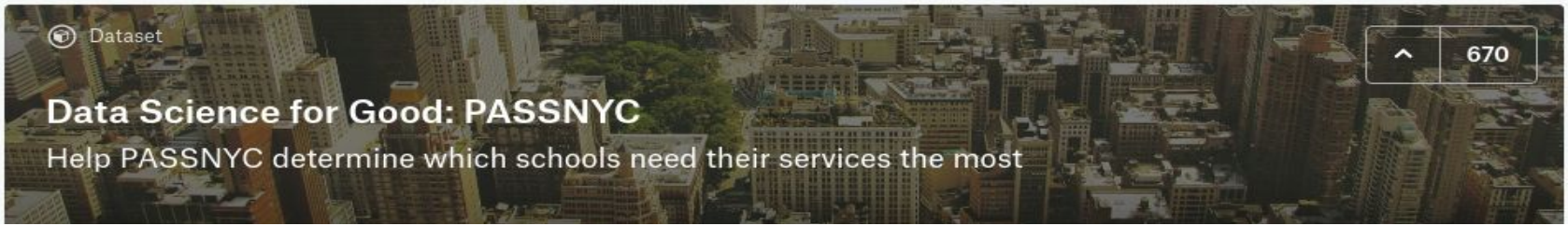
Analytics Competitions:
- ✔ Real data science.
- ✔ Data storytelling.
- ✔ Real Life Solutions
- ✔ Prize Money 💰




INTERNSHALA kaggle LOGICNI Devfolio




INTERNSHALA kaggle LOGICNI Devfolio

Create kernels that impart valuable knowledge.

Share insights that help the data science community.

INTERNSHALA kaggle LOGICNI Devfolio

kaggle DAYS MEETUP

# Why I think kaggle kernels is a semi-successful product?

INTERNSHALA kaggle LOGICNI Devfolio

kaggle DAYS MEETUP

Colab launched around the same time and provides almost all functionalities except version control and “Commit & Chill”.

Colab is mainstream.

# Kaggle has put some restrictions on kernels

Illegal cryptocurrency mining.

Exploitation of hotness algorithm to gain more upvotes

GPU usage is free but not unlimited

INTERNSHALA kaggle LOGICNI Devfolio

kaggle DAYS MEETUP

# The new ASHRAE competition

INTERNSHALA kaggle LOGICNI Devfolio

kaggle DAYS MEETUP

Basically, an American professional organisation of HVAC engineers wants you to predict metered building energy usage in the following areas: chilled water, electric, hot water, and steam meters with millions of rows of training data and even more test data.

kaggle DAYS MEETUP

# THE DATA

| | |
|---|---|
| Training Data | Over 20 million rows |
| Test Data | Over 41.6 million rows |
| Weather data for training set | Around 139,000 rows |
| Weather data for test set | Around 277,000 rows |
| Buildings metadata | Around 1450 rows |

INTERNSHALA kaggle LOGICNI Devfolio

kaggle DAYS MEETUP

# Current trends and tips for the competition

INTERNSHALA kaggle LOGICNI Devfolio

kaggle DAYS MEETUP

# Using Bayesian Optimization

## Why? When? How?

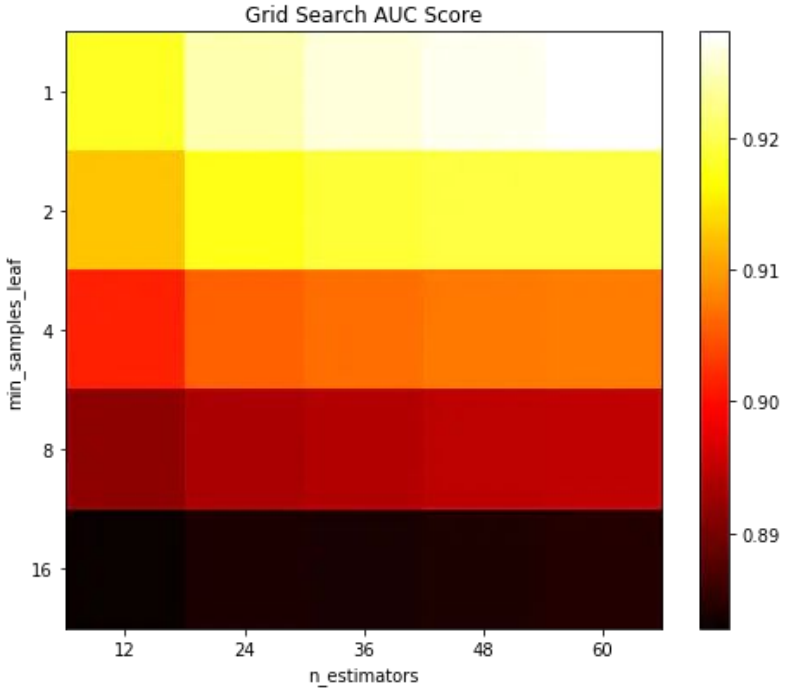
INTERNSHALA kaggle LOGICNI Devfolio

# The Different Methods of Optimization

| | |
|---|---|
| Manual | It is all about guessing and intuition. |
| Random Search | Model is iteratively trained on random parameters by the random search algorithms. |

INTERNSHALA kaggle LOGICMI Devfolio

# The different methods of optimization




## Grid Search

A grid of parameters is formed. Each combination of parameters is used to train models by the grid search algorithm. The best combination of parameters according to the given metric is finally selected.

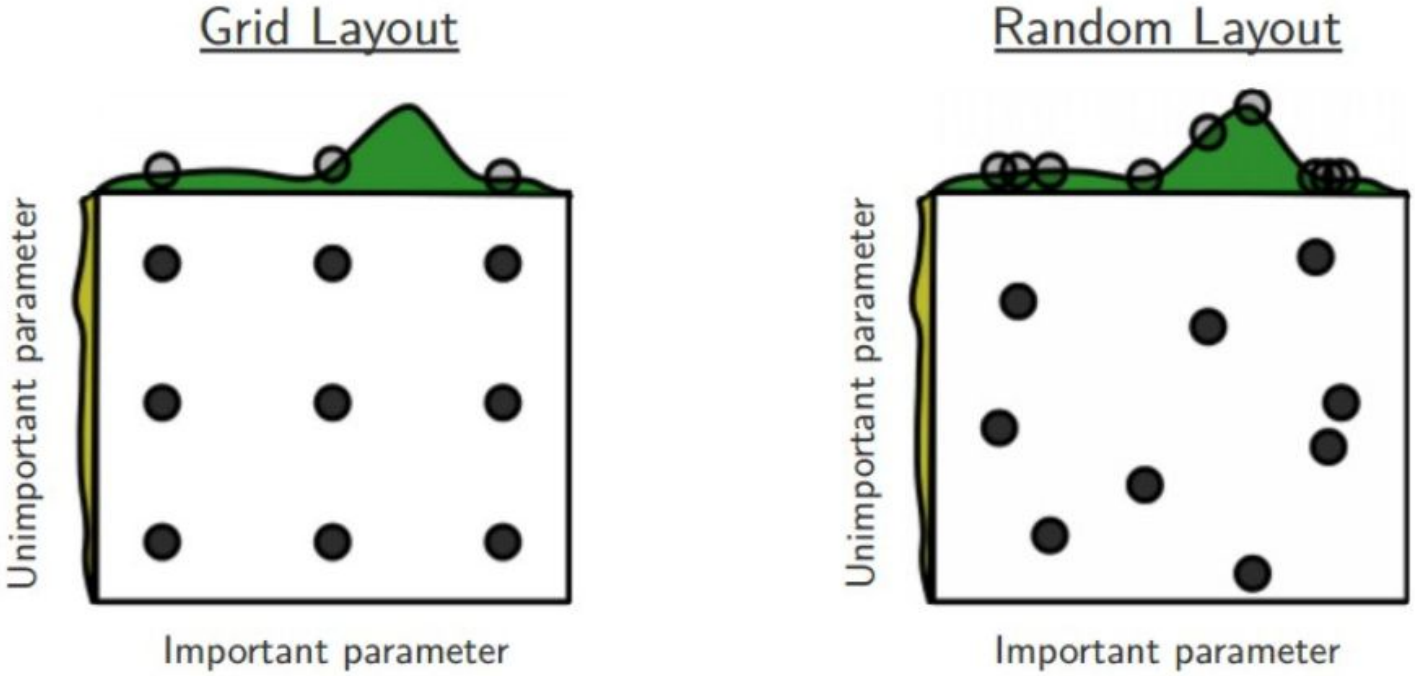



Image from http://jmlr.csail.mit.edu/papers/volume13/bergstra12a/bergstra12a.pdf

INTERNSHALA kaggle LOGICMI Devfolio

```
svc = SVC()
svc.get_params()

param_grid =
[{'C': [0.01,1,10,100,1000],
'kernel':['linear','poly','rbf','s
igmoid'],
'gamma': [0.01,1,10,100,1000]}]
```

kaggle DAYS MEETUP

INTERNSHALA kaggle LOGICNI Devfolio

# The different methods of optimization

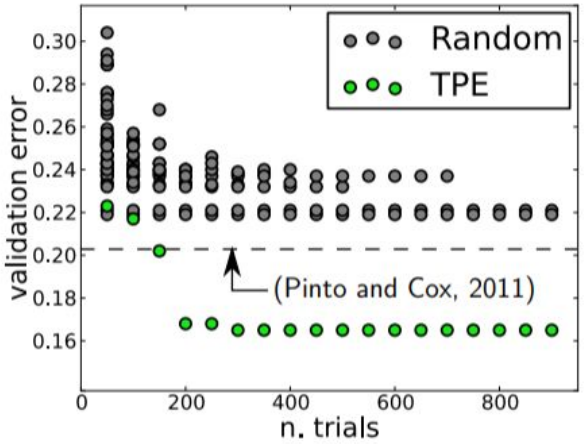
kaggle DAYS MEETUP

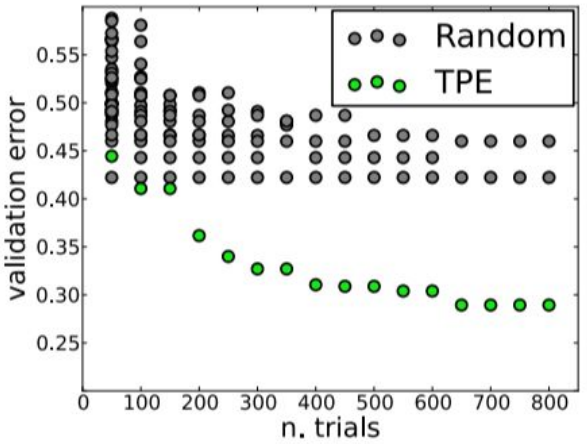
## Bayesian Optimization

The one-sentence summary of Bayesian hyperparameter optimization is: build a probability model of the objective function and use it to select the most promising hyperparameters to evaluate in the true objective function.

Parameters of the next iteration are decided by a probabilistic model which keeps track of past evaluations.

$$P(\,score \mid hyperparameters)$$



# Steps to follow

kaggle DAYS MEETUP

## Define an objective function

This is a function that takes your data, trains the desired model on it and compares the validation set with predictions for the given metric. The metric's value is returned.

# Steps to follow

kaggle DAYS MEETUP

## Define a search space

Usually a dictionary containing either arrays of values for a parameter or a range of values for a parameter ( can be a distribution)

# Steps to follow

kaggle DAYS MEETUP

## Run trials to get best parameters

Run trials/iterations(at least 15-16) using the objective function for the given search space. Eventually the best trial results are given as output.

# When do we use this?

kaggle DAYS MEETUP

Works for large datasets when grid search space is too computationally expensive causing memory errors. High number of trials may not be feasible given large data size.

Surpasses grid search in terms of speed and performance for small datasets and simpler models almost always.

# Results on ASHRAE

kaggle DAYS MEETUP

1.20-1.26 score on various single models

Bayesian Optimization with 20 trials

1.18 on single LightGBM model

Better than single models and performance similar to ensemble model of 15 folds(RMSLE 1.18)

INTERNSHALA kaggle LOGICNI Devfolio

kaggle DAYS MEETUP

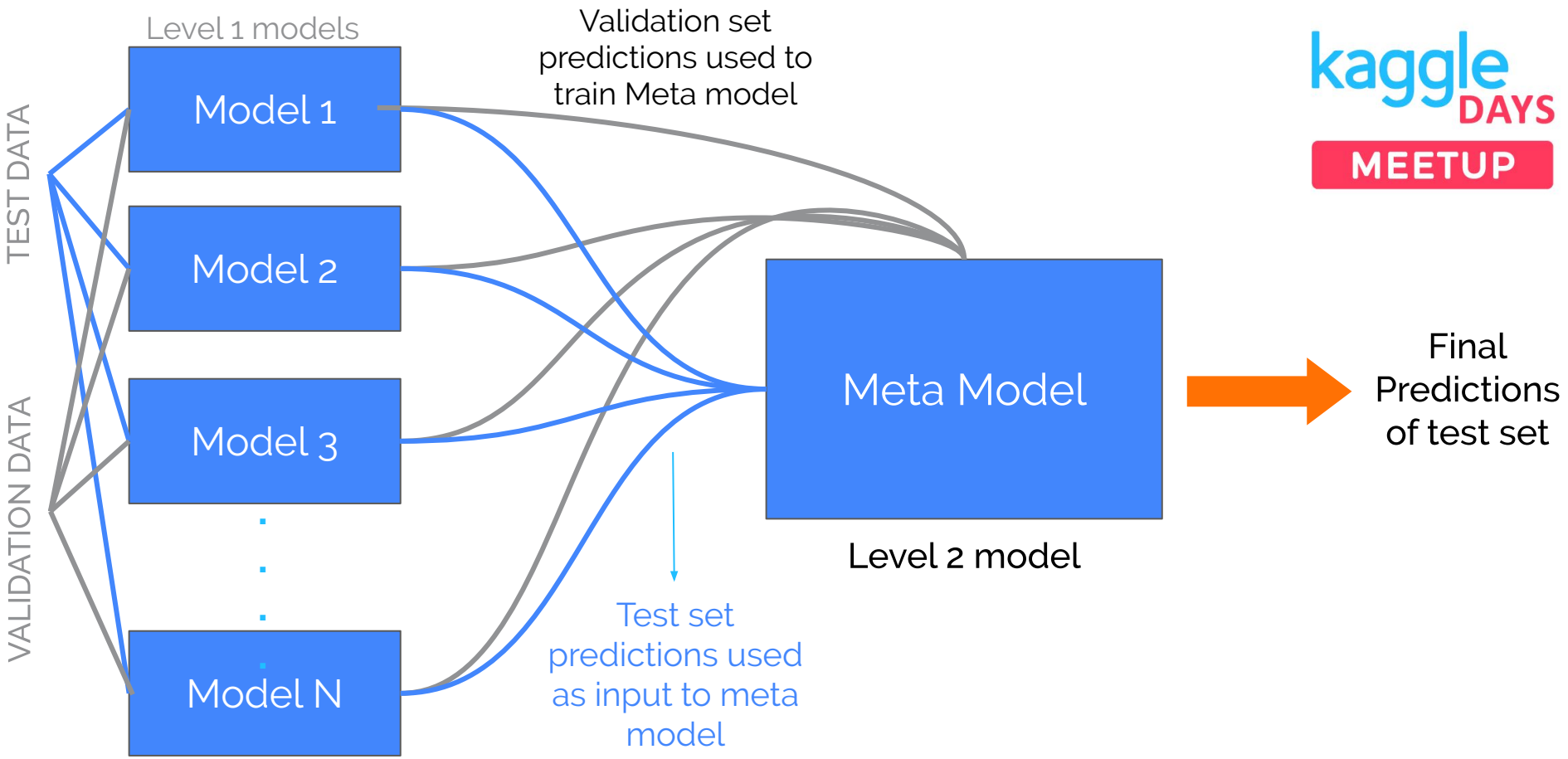
# Stacking

## The rewards of blending

INTERNSHALA kaggle LOGICNI Devfolio

It is an ensembling method where you are using predictions of multiple models on validation set(first level) against real targets to train a new model(second level).

This second level model is then used to get final predictions using first level predictions of test set as its input data.




kaggle DAYS MEETUP

INTERNSHALA kaggle LOGICNI Devfolio

# Results on ASHRAE

kaggle DAYS MEETUP

1.18 using 15 folds LGBM ensemble

1.14 using 10 folds of LGBM+CatBoost with 3-layer Neural Network as Meta Regressor

NN Model is non-deterministic so different predictions each time.

INTERNSHALA kaggle LOGICNI Devfolio

# Inspirations and References


1. https://qiita.com/bmj0114/items/8009f282c99b77780563
2. https://blog.floydhub.com/guide-to-hyperparameters-search-for-deep-learning-models/
3. https://datafai.com/hyper-parameters-optimization-using-gridsearch-and-crossvalidation/
4. https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f

kaggle DAYS MEETUP

# Thanks for your time

INTERNSHALA kaggle LOGICNI Devfolio