

# Demystifying Ensemble Models

Date : 24th May 2020 | Speaker : Ayon Roy | Event : Webinar by Mentorskool

Visit - [AYONROY.ML](http://AYONROY.ML)

# Hello Buddy!

I am **Ayon Roy**

**B.Tech CSE ( 2017-2021 )**

Data Science Intern @ **Lulu International Exchange**, Abu Dhabi  
( **World's Leading Financial Services Company** )

Brought **Kaggle Days Meetup** Community in India for the 1st time

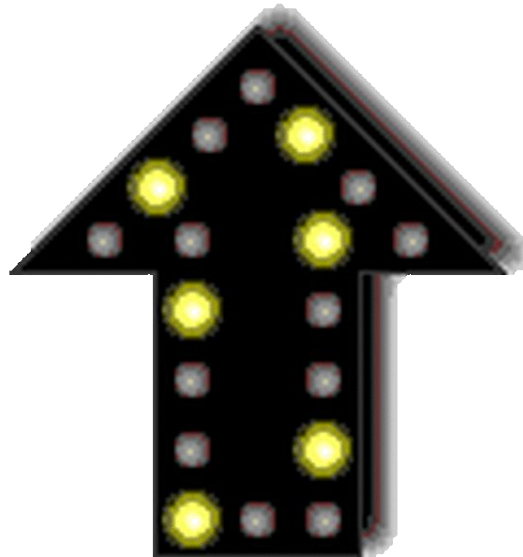
**If you haven't heard about me yet, you might have been living under the rocks. Wake up !!**

# Agenda ( 24-5-2020 )

- What the heck, is Ensemble Models ?
- Why do we need them ?
- What are the Ensembling techniques ?
- Where is it used in Real World ?



# What the heck is, Ensemble Models ?





# Let's assume !!

I want to buy an



Aha !! Buddy, it's month of May

What things will you do ,  
before  
Buying an Air Conditioner ?





Type



Capacity



Power  
consumption



Inverter  
technology



Features

**Room size**

**AC capacity**

Up to 100 sq. ft

**0.8** ton

Up to 150 sq. ft

**1.0** ton

Up to 250 sq. ft

**1.5** ton

Up to 400 sq. ft


**2.0** ton

Visit - [AYONROY.ML](http://AYONROY.ML)

---

## People also ask

---

What does the word ensemble? 

**ensemble.** ... **Ensemble** comes from the Middle French **word** *ensemblée*, which means "together, at the same time." Groups of people who perform at the same time **are ensembles**, so **are** things that **are** put together.

[www.vocabulary.com](http://www.vocabulary.com) › [dictionary](#) › [ensemble](#)

[ensemble - Dictionary Definition : Vocabulary.com](#)

# REAL WORLD DECISION MAKING into MACHINE BASED DECISION MAKING

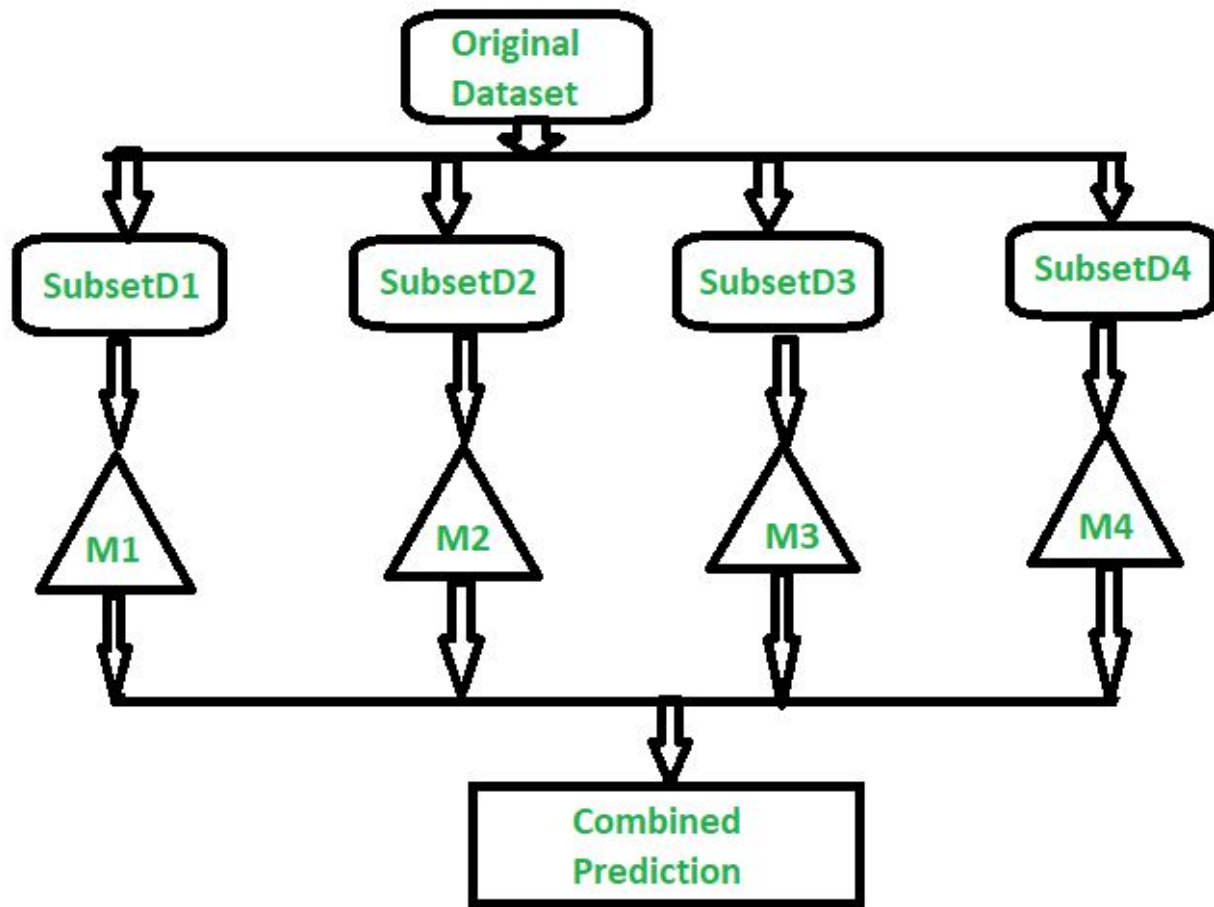


# Ensemble Modelling

Create multiple diverse models to reach a final decision, either by using multiple different modeling algorithms or by using different training data sets.

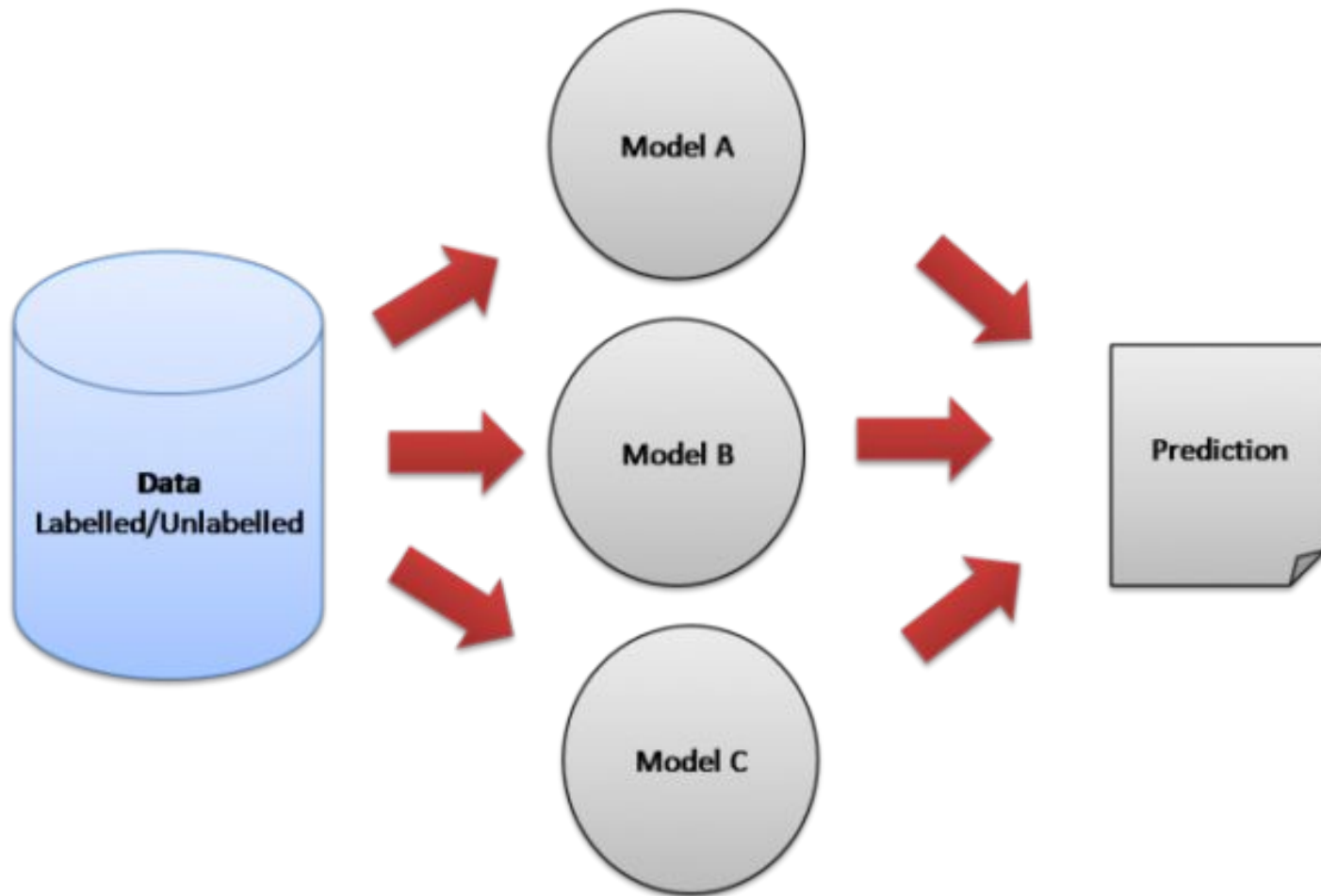
**Ensemble model** then aggregates the prediction of each base model and results in once final prediction for the unseen data.

It is one of the most effective ways to build an efficient machine learning model.



**But**

**Why we need it ?**





**Diversity in making decisions leads to better results as compared to individuals.**

**Similar is true for a diverse set of models in comparison to single models.**

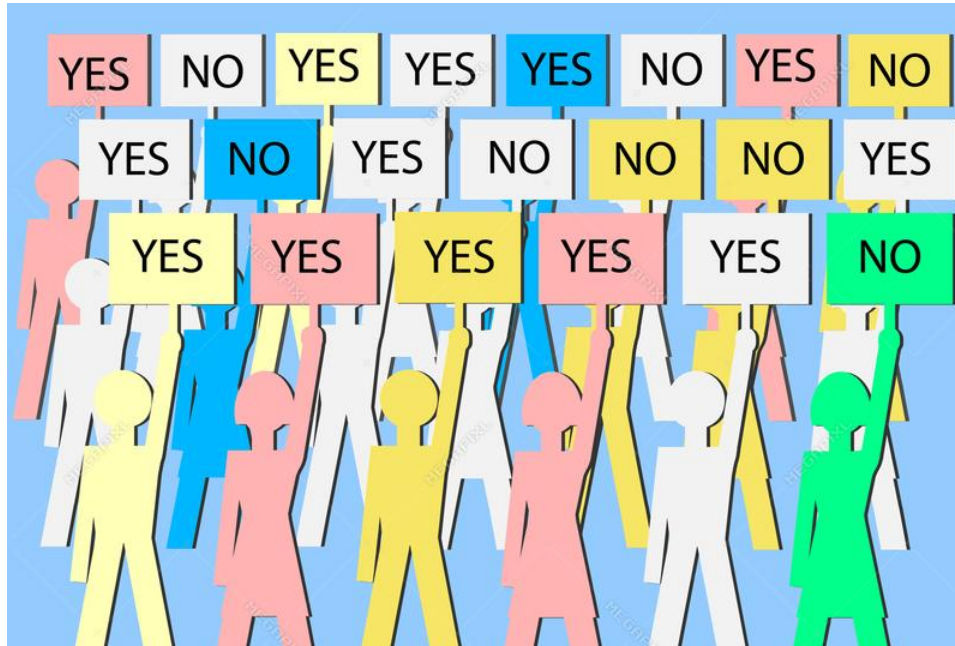




# Basic

# Ensembling Techniques

# Max Voting



**Multiple models are used to make predictions for each data point.**

**The predictions which we get from the majority of the models are used as the final prediction.**

**It is generally used for classification problems.**

# Averaging



**Average  
Formula**

=

**Total Sum of All Numbers**  

---

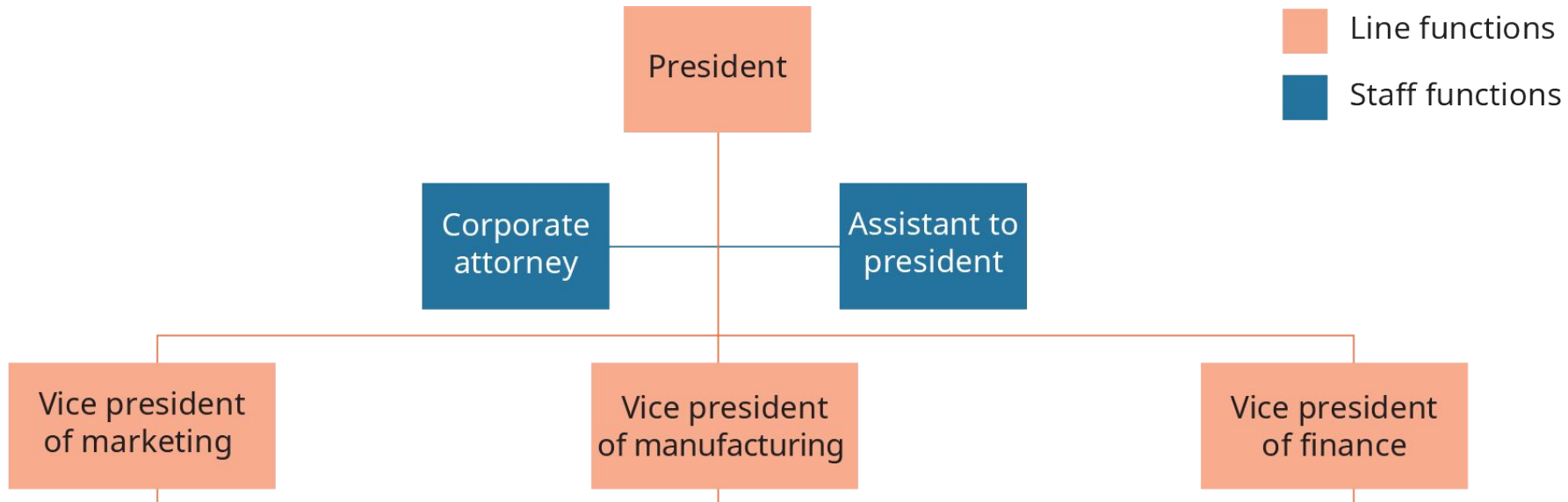
**Number of Item in the Set**



**Average of predictions is taken from all the models and used to make the final prediction.**

**It is used for making predictions in regression problems or while calculating probabilities for classification problems.**

# Weighted Averaging

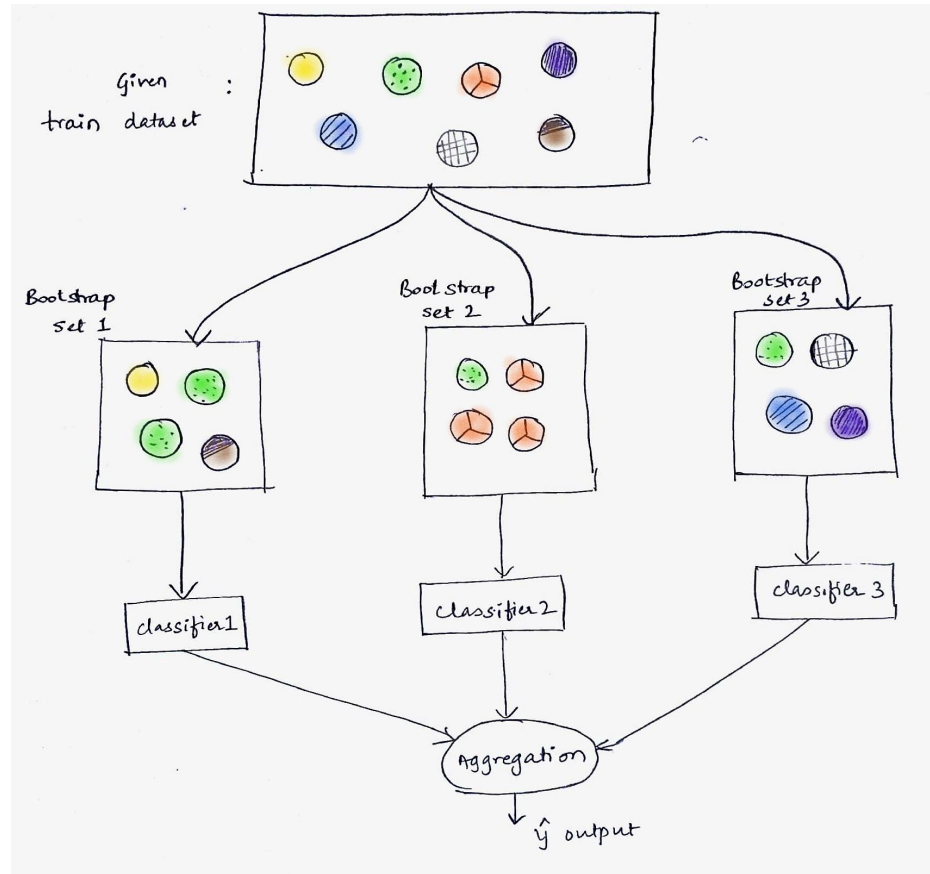


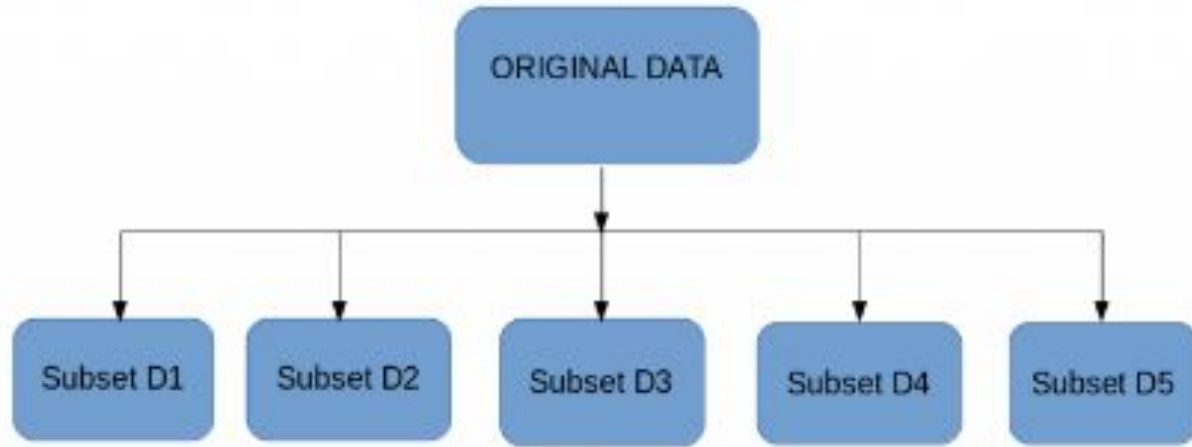
**It is an extension of the averaging method.  
All models are assigned different weights defining the  
importance of each model for prediction.**



# Advanced Ensembling Techniques

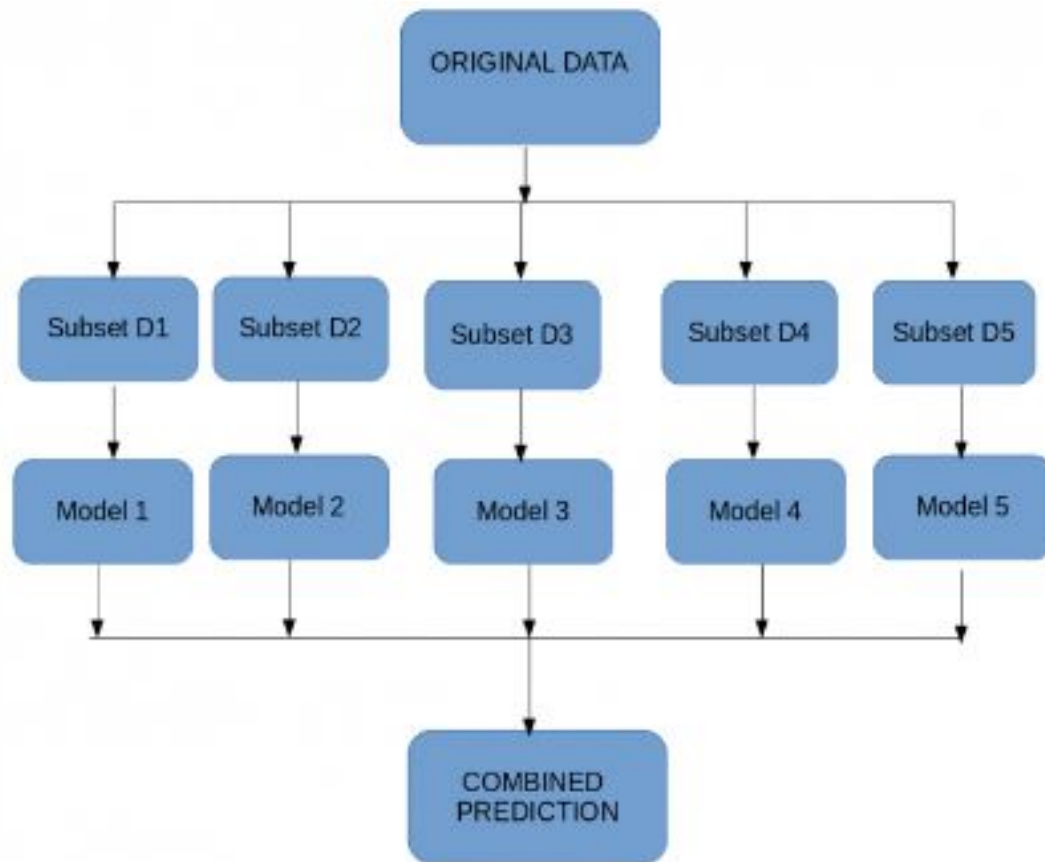
# Bagging ( Bootstrap Aggregating )





**Random sub-samples of the dataset which are drawn from the original dataset randomly with bootstrap sampling method.**

**In bootstrap sampling, some original examples appear more than once and some original examples are not present in the sample.**



# Major Steps

1. Multiple subsets are created from the original dataset, selecting observations with replacement.
2. A base model (weak model) is created on each of these subsets.
3. The models run in parallel and are independent of each other.
4. The final predictions are determined by combining the predictions from all the models.

**Eg : Random Forest use it with some differences**

# Boosting

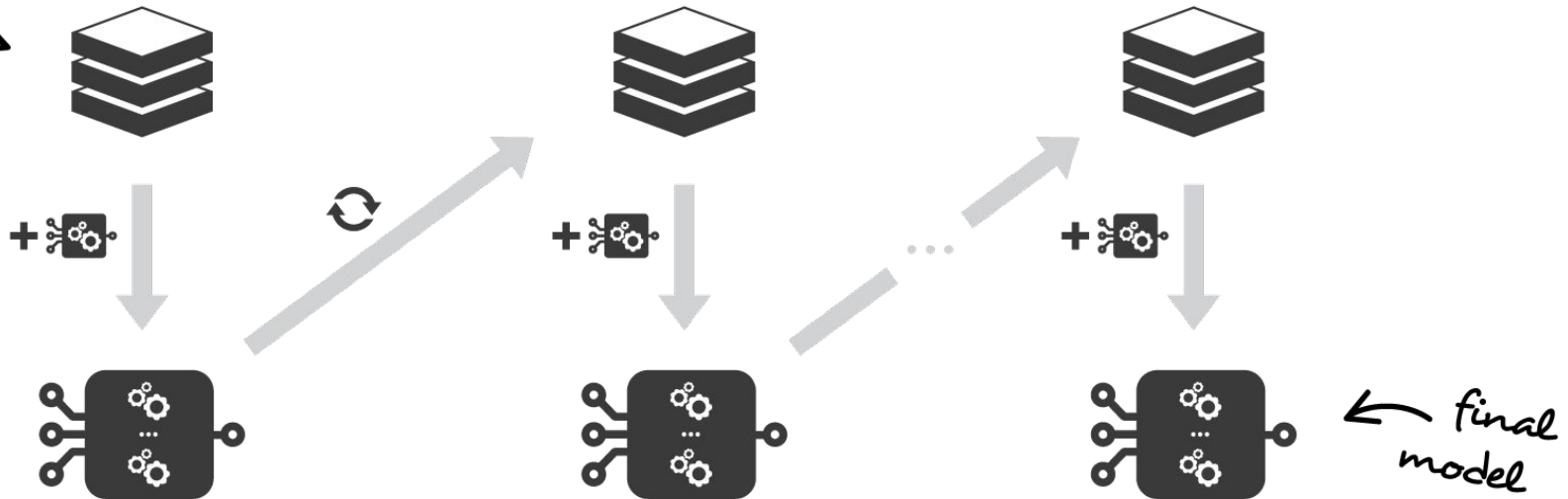


train a weak model  
and aggregate it to  
the ensemble model



update the training dataset  
(values or weights) based on the  
current ensemble model results

*initial  
dataset* →



**A process of converting weak models to strong models.**

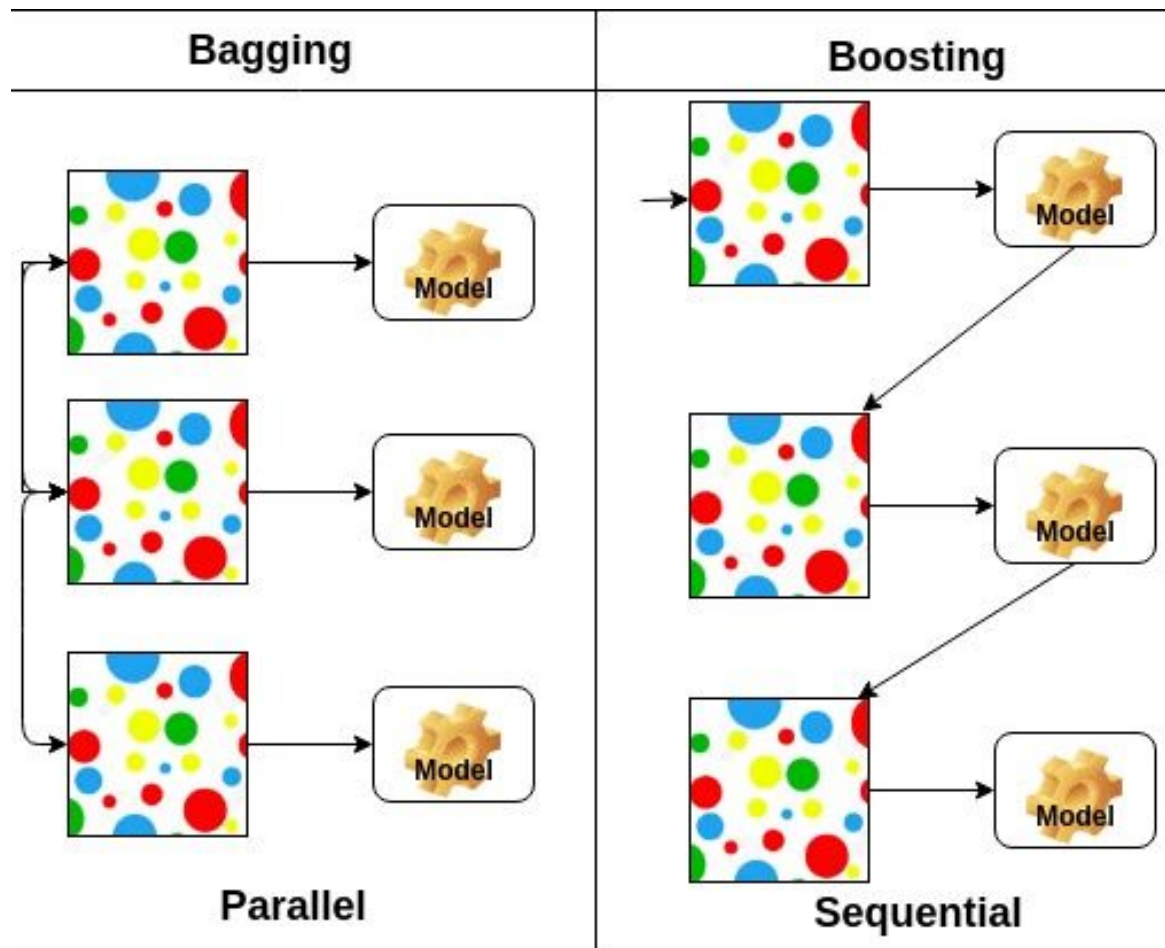
**Boosting incrementally builds an ensemble by training each model with the same dataset but where the weights of instances are adjusted according to the error of the last prediction.**

**Eg: Ada Boost**

# Major Steps

- Draw a random subset of training samples  $d_1$  without replacement from the training set  $D$  to train a weak learner  $C_1$
- Draw second random training subset  $d_2$  without replacement from the training set and add 50 percent of the samples that were previously falsely classified/misclassified to train a weak learner  $C_2$
- Find the training samples  $d_3$  in the training set  $D$  on which  $C_1$  and  $C_2$  disagree to train a third weak learner  $C_3$
- Combine all the weak learners via majority voting.





# And a few more ....

1. Stacking
2. Blending



# Real World Use Case

ZEST  AI

# Performance

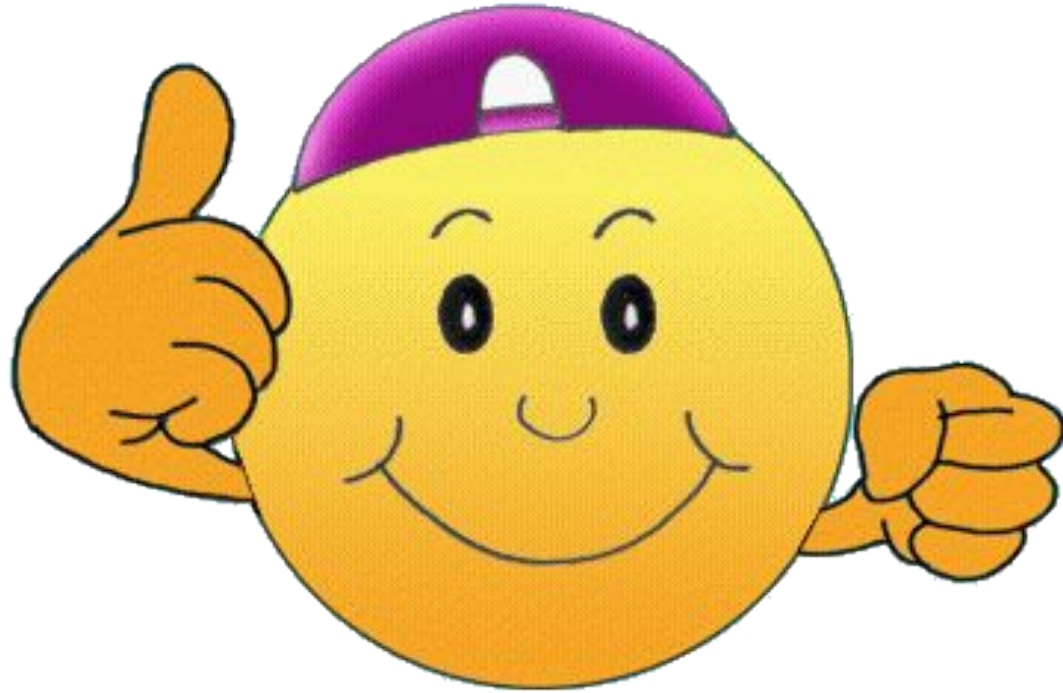
Model Type	AUC	KS	Est. Dollars Saved
Ensemble	0.803	0.446	\$21M
XGB 1	0.791 (2%)	0.420 (6%)	\$18M (14%)
XGB 2	0.791 (2%)	0.428 (4%)	\$18M (12%)
XGB 3	0.781 (3%)	0.411 (9%)	\$17M (16%)
XGB 4	0.782 (3%)	0.413 (8%)	\$17M (16%)
ANN 1	0.750 (7%)	0.376 (19%)	\$16M (19%)
ANN 2	0.786 (2%)	0.430 (4%)	\$18M (13%)

<https://www.zest.ai/article/many-heads-are-better-than-one-making-the-case-for-ensemble-learning>

# Get the resources at

1. <https://medium.com/ml-research-lab/ensemble-learning-the-heart-of-machine-learning-b4f59a5f9777>
2. <https://becominghuman.ai/ensemble-learning-bagging-and-boosting-d20f38be9b1e>
3. <https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725>

**GO FOR IT !**



**GOOD LUCK !**

Let me answer your Questions now.

Finally, it's your time to speak.



# Danke Scheon

Questions ? Any Feedbacks ? Did you like the talk?  
Tell me about it.

If you think I can help you,  
connect with me via

**Email** : [ayonroy2000@gmail.com](mailto:ayonroy2000@gmail.com)

**LinkedIn / Github / Telegram Username** : [ayonroy2000](#)

**Website** : <https://AYONROY.ML/>