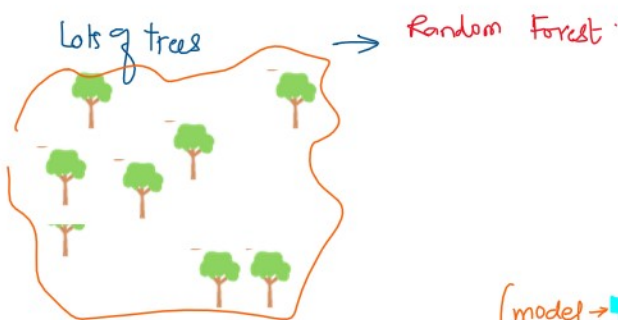




Random Forest



Tree → Decision Tree.



job → (learning)

(model → learner)

Random Forest: It is an ensemble learning algo. that constructs many decision trees during the training.

a technique that combines multiple individual model (DT) to improve the overall predictive performance or stability of the system

more accurate
and robust predictions.

Pros of RF model:

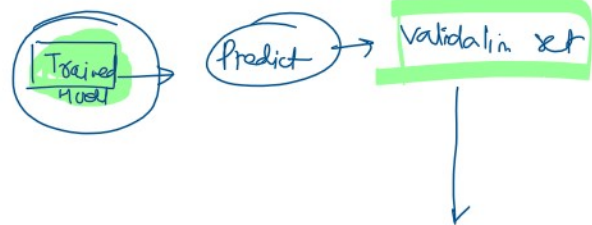
- # Robustness: stable model, less sensitive to noise and overfitting compared to individual DT.
- # High Accuracy: Typically RF model achieves higher accuracy compared to DT.
- # Feature Importance: RF model provides the ranking of features which can help in feature selection and interpretation.

RF can be applied to both regression and classification problems.

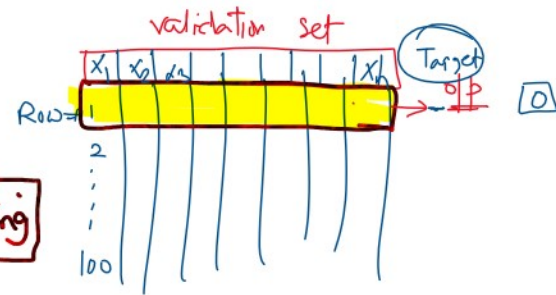
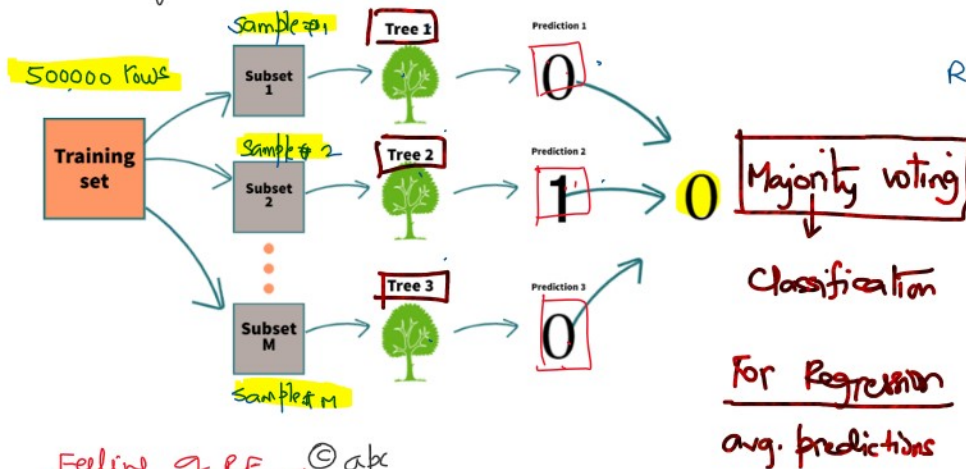
→ Entropy
→ Pruning → max depth.

Cons of Random Forest

- # Model computational time is High.
- # Model interpretability can be less as it has multiple DT models.



Working of RF



Feeling of RF



Disclaimer: Image shown above to understand RF model only.

Every MP in the parliament

DT model

MP - Banglore - subset of India

Bill → Yes → 1
→ No → 0

543 → DT model → MA

Majority

Bagging vs Boosting

RF primarily uses a technique called "Bagging" (aggregating)

↓
stands for
Bootstrap Aggregating

Bagging

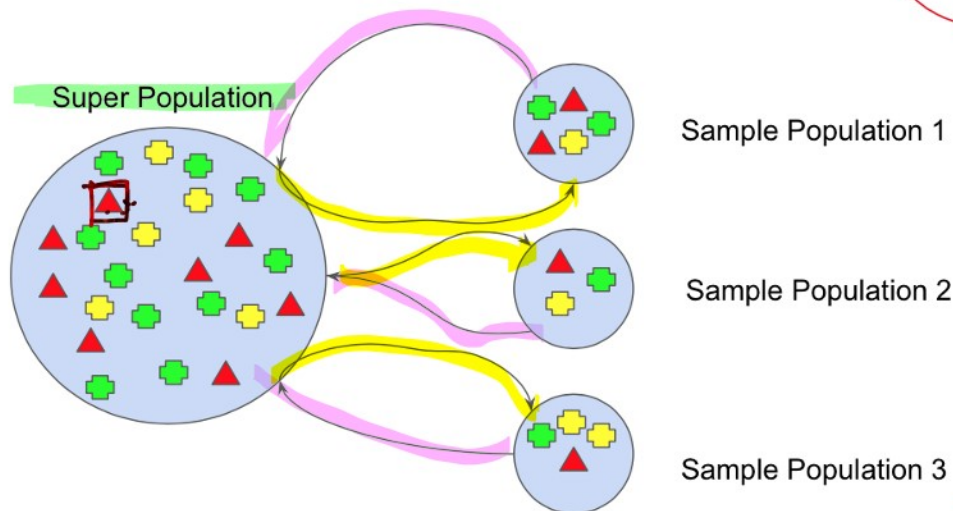
Bagging (Bootstrap aggregation) is a simple and powerful ensemble method which is applied to a high variance machine learning algorithms like decision trees

- Bagging helps to decrease the model's variance.

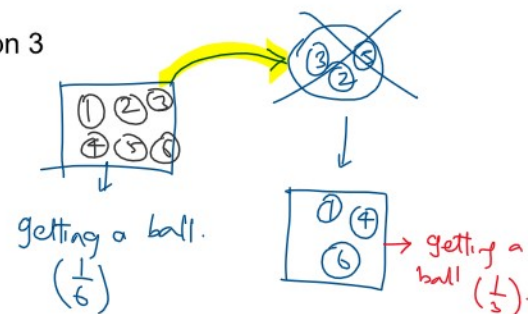
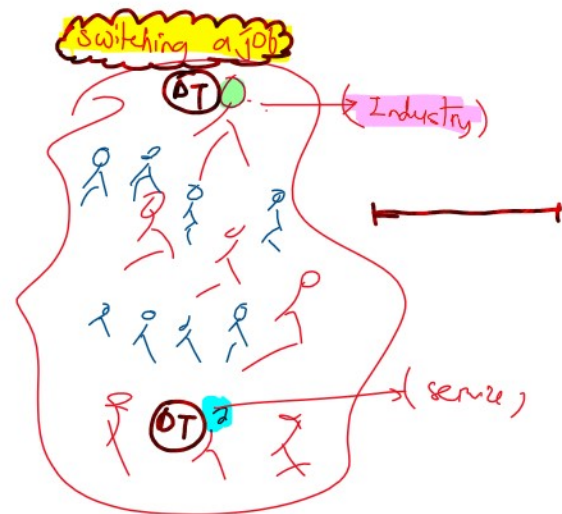
Bootstrap

It refers to random sampling with replacement.

It allows us to better understand the bias and variance within dataset.



- selection of all the datapoints has equal probability.

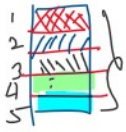


5-fold
↓
Train ... test

K-fold → validation technique.

1

DT models - ⑤



"80% data for training"

K-fold → validation technique

RF → an ensemble ML algo

Random Forest algo always use bagging technique.

association

ML stages

① Load the data

② EDA

③ TRAIN & TEST SPLIT

④ Fit the model into data

Training

⑤ validation

Predict

K-fold
cv

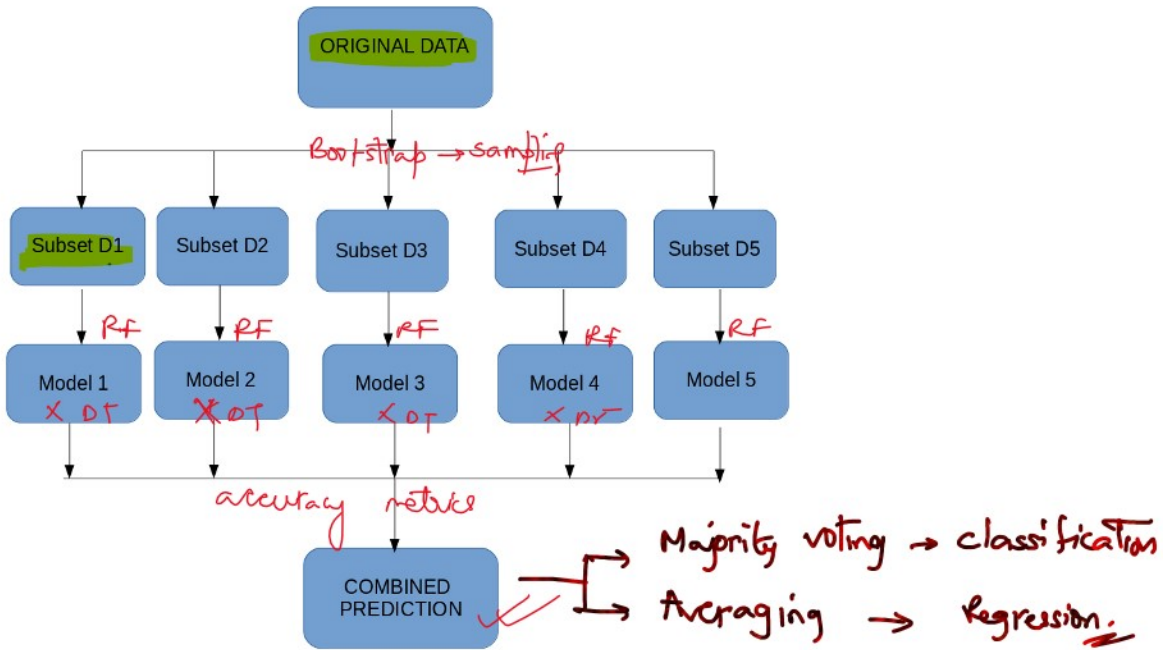
Linear Reg
Log Reg
DT
RF
SVM
KNN

⑥ scores — Acc, F1-Score

Performance
metric

Recall

Accur CF



Boosting

Disclaimer: Image used is to teach boosting (only).



→ (energy drink)

— it reduces the model's bias

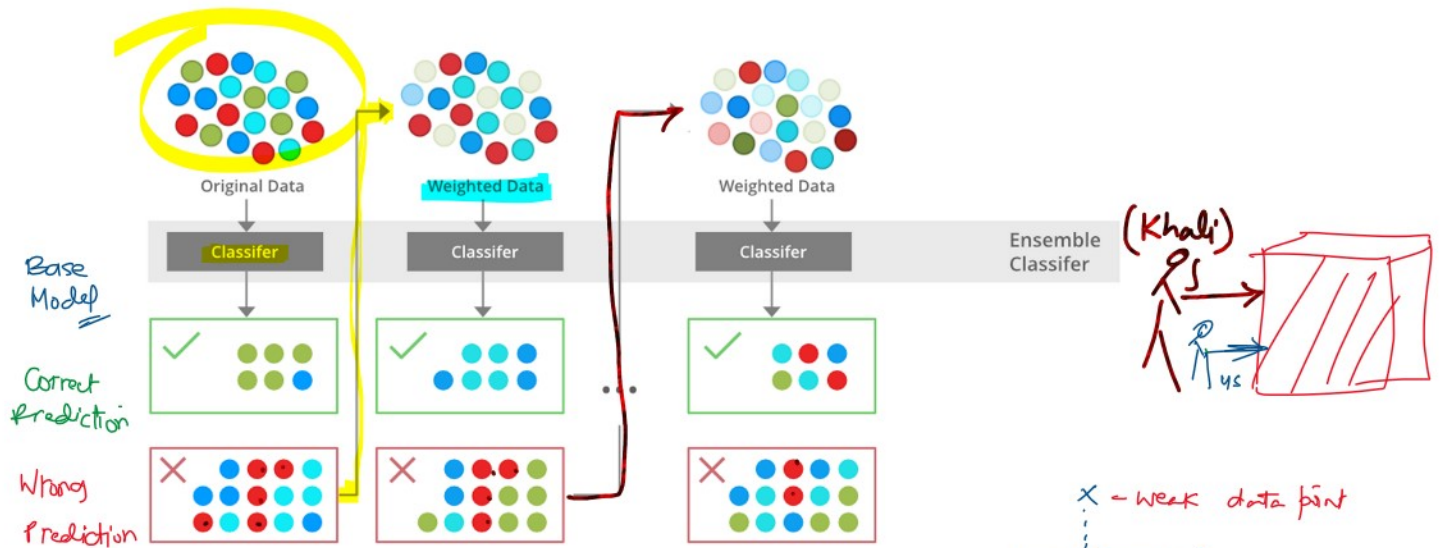
Boosting is a sequential process, where each subsequent model attempts to correct errors of the previous model.

→ Tuition Teacher

↳ Day #01 → 10 prob → mistake in a particular type of question

↳ Day #02 Teacher boosts you with the knowledge

10 prob → ??



- ① A subset is created from the original dataset and initially all data points are given the equal weights
- ② A base model is created on this dataset
- ③ observations which are incorrectly predicted are given higher weights.
- ④ In the next iteration, another model is created and again predictions are made on this dataset
— the next model tries to correct the errors from the previous model.
- ⑤ the final model (strong learner) is the weighted mean of all the models.

GBM: Gradient Boosting Model } Classification & Regression
 XGBM: XGBoost " " " }