Ensembles and its techniques

Analogy
  ↳ DS jobs → Mentor-2
                    ML
              → Mentor-1
  Ajay → data Analytics   Viswasiv ML
                                ↓
                                ML

① data
  ↳ Model 1 → train → Prediction

② data
  { → M1
    → M2
    → M3   } → Combine → prediction
    → M4
    → M5 }
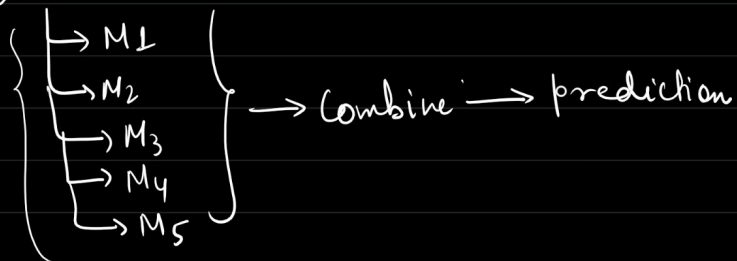
→ One person might give you wrong advice
→ you will connect to multiple mentors
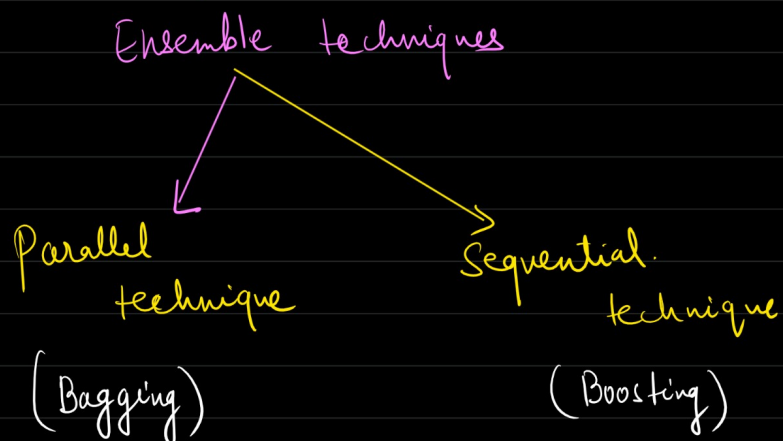→ Chances of getting wrong is minimized when you connect to multiple person.

Ensembles : Combine multiple models
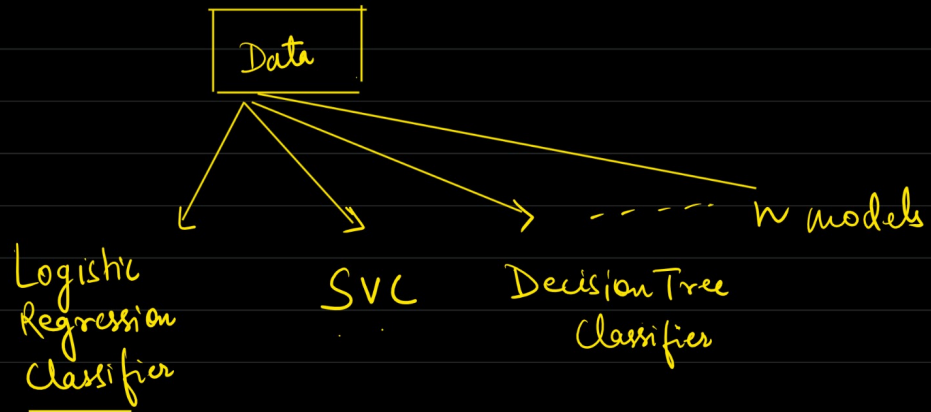 : Prediction — more stable and accurate as compared to Individual model.

(of same Algorithm)
① | DT₁ — (max depth - 5)
  | DT₂ — (max depth - 10)
  | DT₃ — (max depth - 12)

② (different Algorithm)
  → Logist Reg
  → SVC
  → DTC

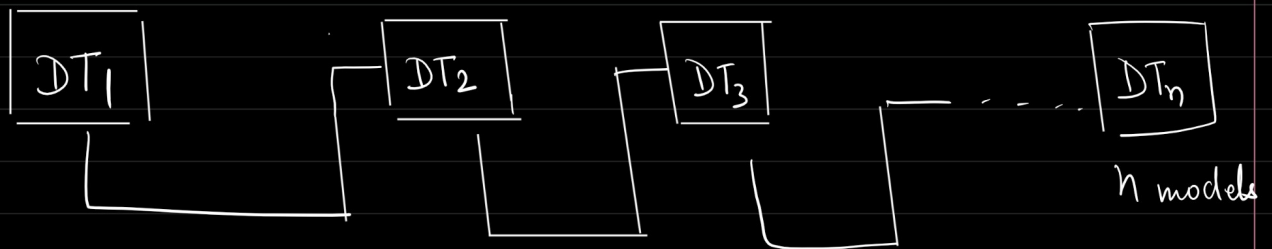* **Ensemble** : Not necessarily only one type of Algorithm.

Ensemble techniques
        ↙   ↘
  Parallel   Sequential.
  technique   technique

  (Bagging)   (Boosting)

Parallel technique of Ensembles

Data

→ Logistic Regression Classifier

→ SVC
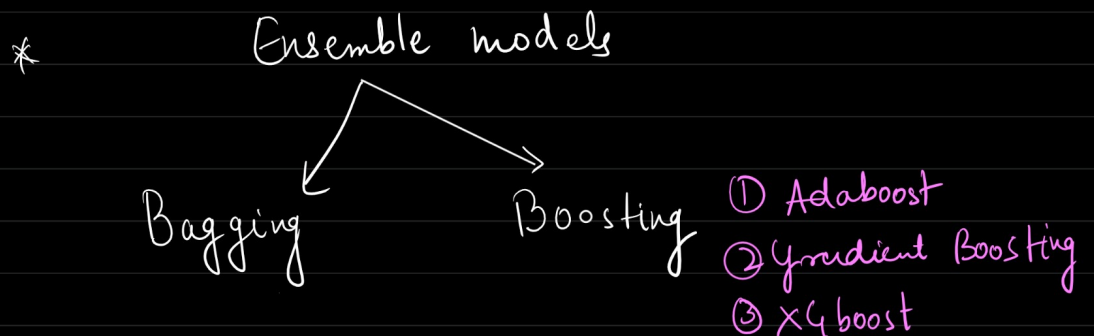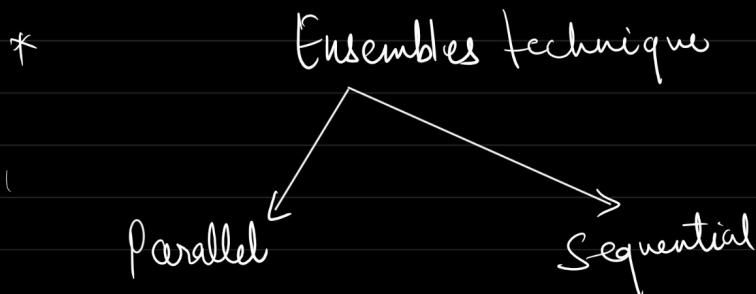
→ Decision Tree Classifier

---- n models

→ All of the models here are built parallely and independent of each other.

* **Sequential technique of Ensemble**

→ All the models are built sequentially and dependent on each other.

DT_1 — DT_2 — DT_3 - - - - DT_n

n models

→ learning from mistake.

* Ensembles technique

Parallel   Sequential

* Ensemble models

Bagging   Boosting   ① Adaboost
                            ② Gradient Boosting
                            ③ XGboost

① Custom Bagging ⟨ Reg / Classifier
② Random forest ⟨ Reg / Classifier

* **Bagging Technique**

→ Parallel models

Bootstrapping
⇓ S < n

(Classification)

Aggregating prediction
⇓ from different base learners

Base learners

pop → {0,2} {3,4} {1,5} {6,4} {3,5}

train data (n)

// Test data //

S' ──── M₁ ──→ Decision Tree classifier ──→ |

S₂ ──── M₂ ──→ Logistic Regression ──→ 0

S₃ ──── M₃ ──→ Support Vector classifier ──→ |

Sₙ ──── Mₙ ──→ N.B ──→ |

Voting Classifier

1
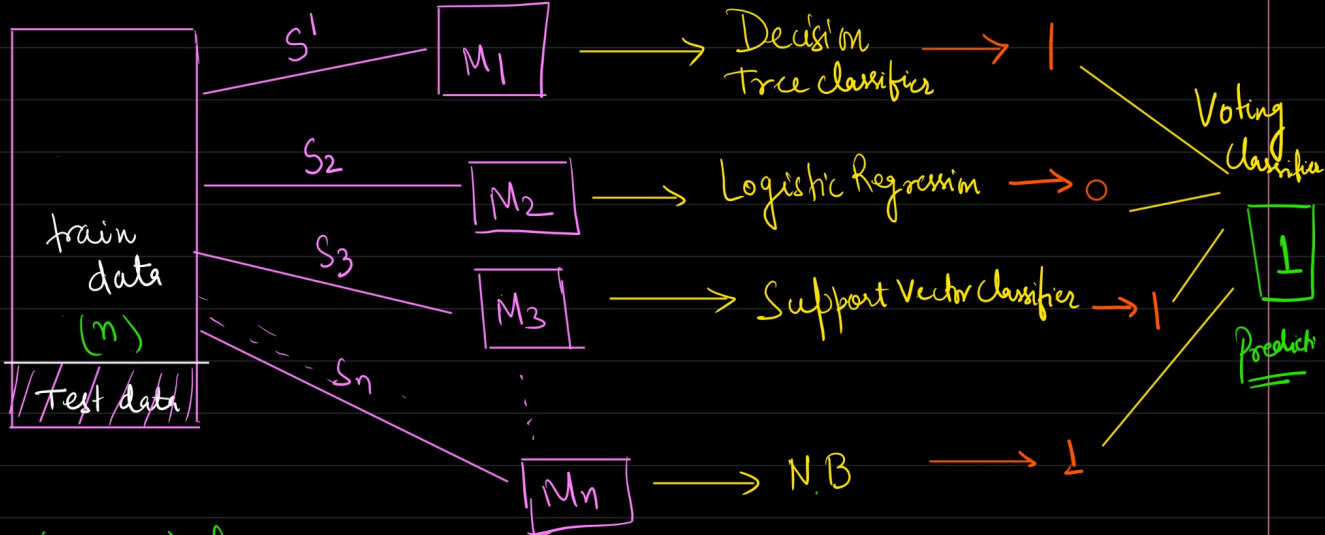
Predict

* Samples (subset) for each model is taken with replacement.

Bagging

Bootstrap ── Aggregating
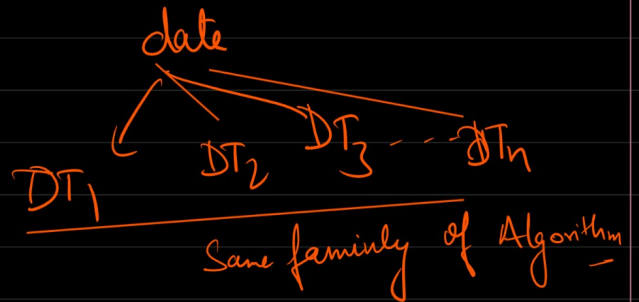
different Samples with replacement
⇓
Boot strap sample

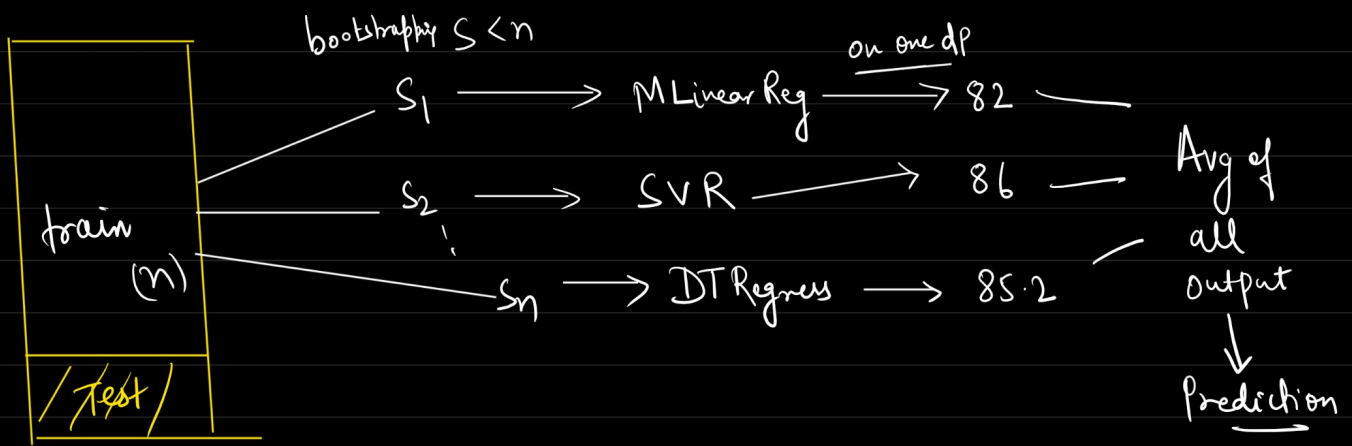Aggregating the prediction.

* **Custom bagging technique**

random forest.

date

{ DT₁   DT₂   DT₃ ---- DTₙ

Same family of Algorithm

Regression

bootstraphy $S < n$

on one dp

$S_1 \longrightarrow$ M Linear Reg $\longrightarrow$ 82

train (n)

$S_2 \longrightarrow$ SVR $\longrightarrow$ 86

$S_n \longrightarrow$ DT Regres $\longrightarrow$ 85.2

Avg of all output

$\downarrow$

Prediction

/ Test /

$\longrightarrow$ Custom Bagging Regressor | classifier.

$\longrightarrow$ Pipeline | Column Transformer (missing value, Encoding, Scaling)

* Random Forest