Bagging And Boosting
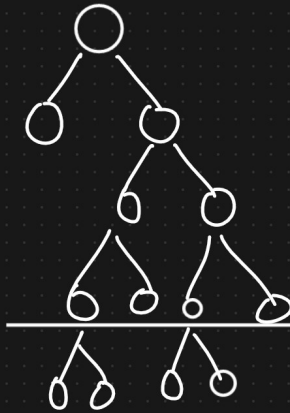
① Random Forest Classifier
② Random Forest Regressor

Decision Tree

Overfitting

Low Bias
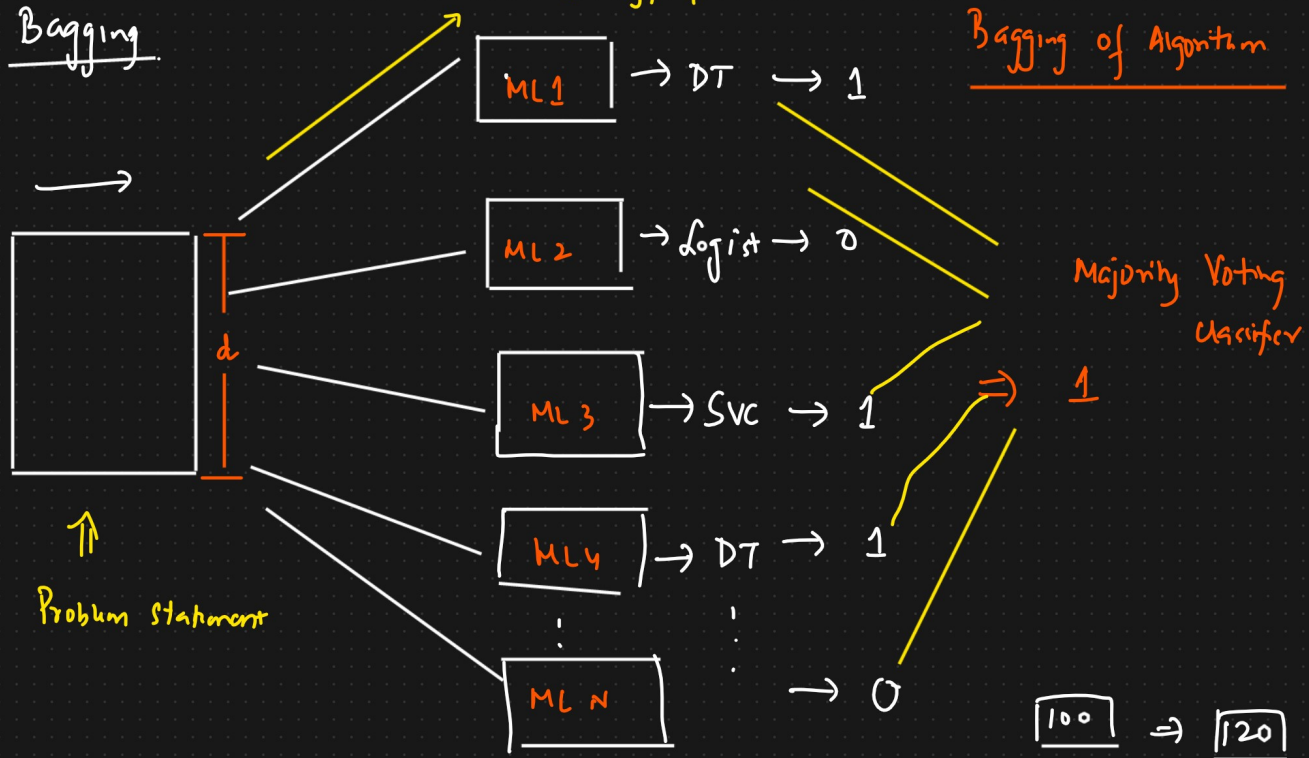High Variance
$\Rightarrow$
↑
Prepruning
Post prunning

Bagging And Boosting

[Ensemble Techniques]
Bootstrap Aggregation

① **Bagging**

$\rightarrow$

Bagging of Algorithm

ML1 $\rightarrow$ DT $\rightarrow$ 1

ML2 $\rightarrow$ Logist $\rightarrow$ 0

ML3 $\rightarrow$ Svc $\rightarrow$ 1

ML4 $\rightarrow$ DT $\rightarrow$ 1

ML N $\rightarrow$ 0

d

↑
Problem Statement

Majority Voting
Classifier

$\Rightarrow$ 1

100 $\Rightarrow$ 120

KBC : Kaun Banega Crorepati [Amitabh Bachan]

⎰ KRISH $\longrightarrow$ DATA SCIENCE
⎱ UPSC $\longrightarrow$ Diversified

↓ ↓ ↓ ↓ ↓ ↓ ↓
□ □ □ □ □ □ ⊓
↑ ↑ ↑ ↑ ↑ ↑

KBC $\longrightarrow$

$\rightarrow$ Current Affairs $\longrightarrow$ A

$\Rightarrow$ Economics $\longrightarrow$ A

$\Rightarrow$ History $\longrightarrow$ B

$\Rightarrow$ Science. $\longrightarrow$ A

Majority Voting

$\Rightarrow$ A

Average or

Mean

Ensemble Techniques
$\Downarrow$

Kaggle, Hacker Rank
$\Downarrow$

Outperforming results

Bagging

① Random Forest Classifier

② Random Forest Regression

② **Boosting**

Strong Learner

Weak Learners
$\Downarrow$

WL   WL $\downarrow$ WL

M1 $\longrightarrow$ M2 $\longrightarrow$ M3 $\rightarrow \rightarrow \rightarrow$ Mn

① AdaBoost Regressor And Classifier

② Gradient Boost   "   ^   "

③ Xgboost   "   "   "

$\hookrightarrow$ Extreme Gradient Boost
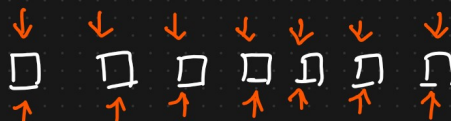
Random Forest CLASSIFICATION And Regression $\Rightarrow$ Ensemble Technique

$\Downarrow$

English

Bagging And Boosting

Combine Multiple DT's

d'<d

h'<m

m

Row sampling (d')
+ (m')
Feature sampling + with Replacement

|200|

M1 $\rightarrow$ DT1 $\rightarrow$

Row Sampling
Feature + Sampling

|200|

M2 $\rightarrow$ DT2 $\rightarrow$

1000

d

2o.

M3 $\rightarrow$ DT3 $\rightarrow$

200

Average $\Rightarrow$ Generalized Models

200.

T

2o

Validation

$\rightarrow$ Validation data.

Mn $\rightarrow$ DTn $\rightarrow$

New Test Data

Bootstrap

Aggregation
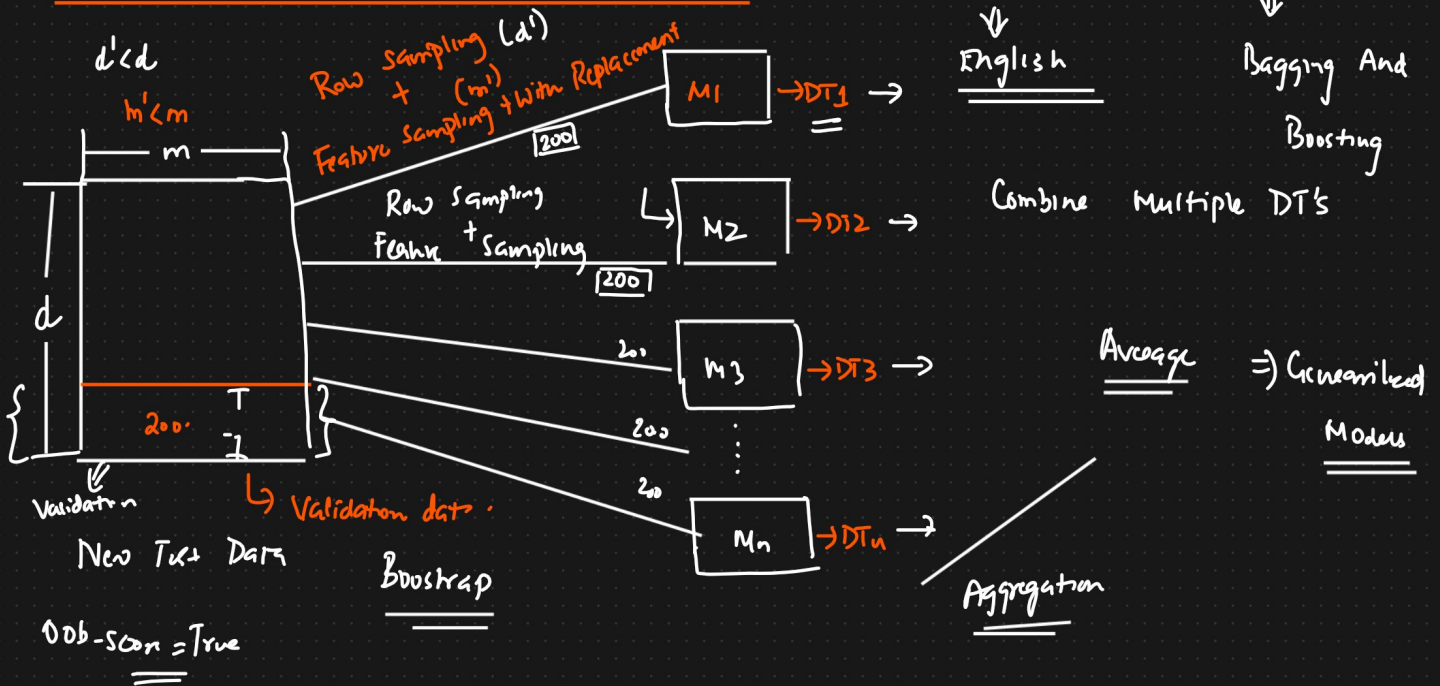
OOb-score =True

DT1 $\Rightarrow$ Overfitting $\Rightarrow$ Prepruning or Post Prepruning $\Rightarrow$ Cubersome Task

$\checkmark$ ① Low Bias $\longrightarrow$ Training

$\checkmark$ ② High Variance $\longrightarrow$ Low Variance

$OOB\_{score} = TRUE$

DATASET

SKlearn

TRAIN

Test

Out of Bag Evaluation

TRAIN

Validaton

CROSS VALIDATION
GRID SEARCHYU

Model Tracing

Hyperparameter Tunning

$\Downarrow$
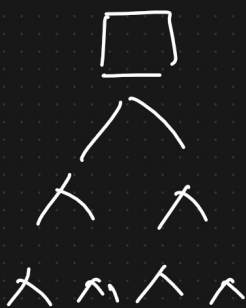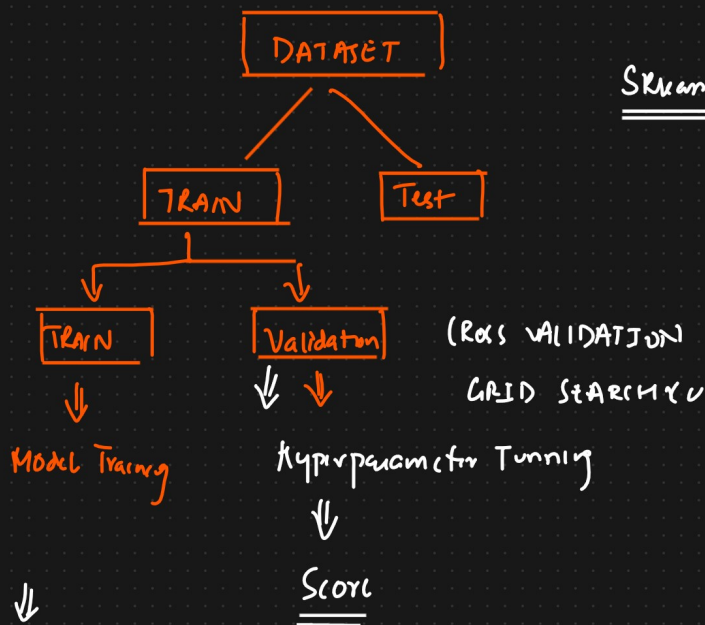
Score

$\boxed{Oob\_score} = \boxed{83\%}$

$\Downarrow$

Accuracy   Validation

Out of Bag Error

$1 - Oob\_score = 1 - 83 = 0.17\%$ {Validation Error}