Start-Tech Academy

Random Forest
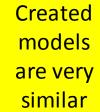
Problem:-
Bagging creates correlated trees

**Shortcomings
Of Bagging**

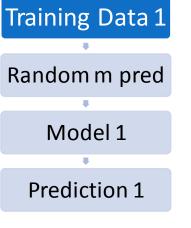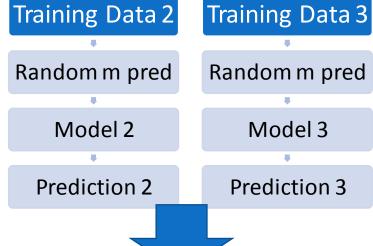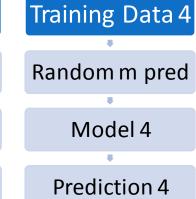Created models are very similar

Final Model

Random Forest

Concept:-
We use subset of predictor variables so that we get different splits in each model

Shortcomings

Different set of m predictors out of p

Training Data 1	Training Data 2	Training Data 3	Training Data 4
Random m pred	Random m pred	Random m pred	Random m pred
Model 1	Model 2	Model 3	Model 4
Prediction 1	Prediction 2	Prediction 3	Prediction 4

Final Model

Random Forest

Thumb Rule for value of M

1. For Regression
   P/3

1. For Classification
   sqrt(P)

2. Don't forget to use your business knowledge
   If the variables are highly correlated try a smaller value of M