1. 3 stages of model in Machine Learning

->**Training** the model using Train data from the known population

->**Testing** the model using Test data from the known population

->**Validation** of the model to tune the performance of the model and choosing the model

1. Standard approach for Supervised Learning

a) Problem Statement definition

b) Data Sourcing from the right source

c) Exploratory Data analysis (EDA) to understand the data

d) Feature engineering and Data transformation (Pre processing)

e) Data Splitting (Train ,Test, Validate)

f) choosing model and Model creation with Train data

g) Testing the model with test data

h) Validating the model and fine tuning

i) Deploying the model in to production

1. Training data set and Test data

Training data set: Data used to train and develop the model

Test Date set: Data used to test the accuracy and performance of the model

1. Ensemble model: A final model created by using multiple models using single or multiple algorithms so the performance can be better than single model

Bagging: Several Subsets of data are creates using random sample with replacement. Models are created using multiple data sets.Average of the models result is taken

Boosting: In boosting technique learning is performed sequentially .A model learns from its previous model and rectifies the errors. The net error by the final model will be less.Group of weak learners used to build a strong learner

5)Overfitting: Overfitting can be handled by reducing R2(Coefficient of determination) and splitting the data randomly using 70 30 rule will help to reduce over fitting.Bagging technique also reduces the overfitting.