1. **What are the three stages to build the hypotheses or model in machine learning?**

**Ans:**

1. Model building
2. Model evaluating
3. Model Applying

**2) What is the standard approach to supervised learning?**

**Ans:**

1. EDA
2. Data Visualisation
3. Handling null values
4. Train/ Test split
5. Training the model using train data
6. Evaluating model using test data
7. Further tuning to improve the efficiency
8. Implementation of model

**3) What is Training set and Test set?**

**Ans:**

**Training Set:** Data set to train the model. Model is trained based on the features and labels of the Training set. Both features and labels are shown to the model. Normally 70% of dataset is given for training.

**Test Set:** Dataset used to evaluate the trained model. Only the features are given to the model and the model gives the labels. Test label and output of the model are compared and efficiency of the model is calculated. Normally 30% of dataset is given for testing.

**4) What is the general principle of an ensemble method and what is bagging and boosting in ensemble method?**

**Ans:**

**Ensemble Method**: Ensemble method is combining different models to get higher accuracy and precision. Ensemble method has two types 1) Bagging and Boosting.

**Bagging**: Bagging is group of same models. Dataset is given to different models (sample with replacement – bootstrap) and the output of the each model is combined. It normally decreases the variance without increasing the bias.

**Boosting**: Boosting is like mismatch in each model is given as the input to the next model . At last, error will be minimized. It normally decreases the bias without increasing the variance.

**5) How can you avoid overfitting?**

**Ans:**

1. Using more number of data
2. Implementing Cross Validation