# Ensemble Methods

1) **What is Ensemble Models?**
   Ensemble models improve model accuracy by combining the results from many models.
2) **What is bagging?**
   Bagging is a type of ensemble model based on fitting many models to bootstrapped samples of the data and averaging the models.
3) **What is Random Forest?**
   Random Forest is a special type of bagging applied to decision trees. In addition to resampling the data, the random forest algorithm samples the predictor variables when splitting the trees.
4) **Why Random Forest is useful?**
   • A useful output from the random forest is a measure of variable importance that ranks the predictors in terms of their contribution to model accuracy.
   • The Random Forest has a set of hyper parameters that should be tuned using cross-validation to avoid overfitting.
5) **What is Variable Importance?**
   A measure of the importance of a predictor variable in the performance of the model.
6) **What is Boosting?**
   A general technique to fit a sequence of models by giving more weight to the records with large residuals for each successive round.
7) **What is Adaboost?**
   An early version of boosting based on reweighting the data based on the residuals.
8) **What is Gradient Boosting?**
   A more general form of boosting that is cast in terms of minimizing a cost function.
9) **What is Stochastic Gradient Boosting?**
   The most general algorithm for boosting that incorporates resampling of records and columns in each round.
10) **What is Regularization?**
    A technique to avoid overfitting by adding a penalty term to the cost function on the number of parameters in the model.
11) **What is Hyper-parameters?**
    Parameters that need to be set before fitting the algorithm
12) **What is cross-validation?**
    Cross validation is especially important for boosting due to the large number of hyper parameters that need to be set.
13) **What is stacking?**
    In stacking, initially you train multiple base models of different types on training/testing datasets. It is ideal to mix models that work differently.
14) **What is hard voting and soft voting?**
    Majority voting is also known as hard Voting.
    The argmax of the sum of predicted probabilities is known as soft voting.