1. Is there any way to combine five different models that have all been trained on the same training

data and have all achieved 95 percent precision? If so, how can you go about doing it? If not, what is

the reason?

Creating a train and test split of your dataset is one method to quickly evaluate the performance of an algorithm on your problem.The training dataset is used to prepare a model, to train it.

We pretend the test dataset is new data where the output values are withheld from the algorithm. We gather predictions from the trained model on the inputs from the test dataset and compare them to the withheld output values of the test set.

Comparing the predictions and withheld outputs on the test dataset allows us to compute a performance measure for the model on the test dataset. This is an estimate of the skill of the algorithm trained on the problem when making predictions on unseen data.

2. Whats the difference between hard voting classifiers and soft voting classifiers?

In classification problems, there are two types of voting: hard voting and soft voting. Hard voting entails picking the prediction with the highest number of votes, whereas soft voting entails combining the probabilities of each prediction in each model and picking the prediction with the highest total probability.

3. Is it possible to distribute a bagging ensembles training through several servers to speed up the

process? Pasting ensembles, boosting ensembles, Random Forests, and stacking ensembles are all

options.

It is quite possible to speed up training of a bagging ensemble, pasting ensembles and Random Forests by distributing it across multiple servers, since each predictor in the ensemble is independent of the others.

4. What is the advantage of evaluating out of the bag?

Random forests do not require a validation dataset. Most random forests use a technique called out-of-bag-evaluation (OOB evaluation) to evaluate the quality of the model. OOB evaluation treats the training set as if it were on the test set of a cross-validation.

5. What distinguishes Extra-Trees from ordinary Random Forests? What good would this extra

randomness do? Is it true that Extra-Tree Random Forests are slower or faster than normal Random

Forests?

Random Forest chooses the optimum split while Extra Trees chooses it randomly. However, once the split points are selected, the two algorithms choose the best one between all the subset of features. Therefore, Extra Trees adds randomization but still has optimization

6. Which hyperparameters and how do you tweak if your AdaBoost ensemble underfits the training

data?

If your adaboost ensemble underfits the training data, which hyperparameters should you tweak and how? You can try increasing the number of estimators or reducing the regularization hyperparameters of the base estimator. You may also try slightly increasing the learning rate.

7. Should you raise or decrease the learning rate if your Gradient Boosting ensemble overfits the

training set?

If your gradient boosting ensemble overfits the training set, you should try decreasing the learning rate.