***Ans 1***

1. Model Building
2. Model Testing
3. Applying the Model

**Ans2**

Split sample data into a training set and test set

**Ans3**

In machine learning, a training set is a dataset used to train a model. In training the model, specific features are picked out from the training set. These features are then incorporated into the model. Thereby, if the training set is labeled correctly, the model should be able to learn something from these features.

The test set is a dataset used to measure how well the model performs at making predictions on that test set. If the prediction scores for the test set are unreasonable, we’ll need to make some adjustments to our model and try again.

The issue here is that our test would yield misleading results if we test our model with the training data. The model itself was created by learning from the training set, so it will likely do quite well at making predictions on the training set itself- it knows this data too well. We need to test the model with a test set, i.e. a dataset the model hasn’t seen before.

**Ans 4**

Combine predictions of several models in order to improve robustness over a single model... Bagging is used to improve unstable estimation and classification, while boosting is used to reduce bias in an ensemble.

**Ans 5**

Decision trees are a nonparametric machine-learning algorithm that is very flexible and is subject to over fitting training data. This problem can be addressed by pruning a tree after it has learned in order to remove some of the detail it has picked up.