Executive Summary:

The objective of this machine learning exercise is to create a prediction model for the 'Classe' variable using a combination of the other variables.

Analysis was performed using two types of algorithms – Decision Tree and Random Forests. The Random Forests approach provided better results with almost perfect sensitivity, specificity, positive and negative predictive value.

Thus, the Random Forests model is used as the model of preference for this classification exercise.

Details:

The following pre-steps were performed towards this analysis.

1. The training and testing data sets were retrieved from the given URL.
2. Data cleansing
   a. The identification was removed to nullify it's impact on the model
   b. Variables with no variation were removed

Thereafter, the training data was used to create two models,

1. Decision Tree based Model
2. Random Forest based Model

The models were trained using the training data sets.

Thereafter the models were tested for the following metrics using a confusionMatrix.

- Sensitivity
- Specificity
- Positive predictive value
- Negative predictive value

Based on the values of the above metrics, the 'Random Forest' model is deemed more appropriate and accurate for prediction.

Sample Procedure:

Getting in the training set –

> trainingData<-read.csv("pml-training.csv",header=TRUE)

> testingData<-read.csv("pml-testing.csv",header=TRUE)

These were massaged to create a refined training data set 'newTrainingData' and a refined testing set 'newTestingData'

Method 1: Decision Tree

The following step was done to create a decision tree based prediction model.

> modelDecTree<-rpart(classe~.,data=newTrainingData,method="class")

The same was then used to perform prediction.

> newPrediction <- predict(modelDecTree, newTestingData, type = "class")

Method 2: Random Forest

The following step was done to create a random forest based prediction model.

> modelRandomForest<-randomForest(classe~.,data=newTrainingData)

The same was then used to perform prediction.

> newPrediction <- predict(modelRandomForest, newTestingData, type = "class")

Result:

The random forest approach yielded much better prediction results and would be the preferred model for predicting 'classe'.