The problem to be modeled is essentially a classification problem and we define certain performance assessment criteria to evaluate the performance of each applied algorithm and derive inferences from it. The evaluation metrics for our models are listed as follows

5.1.1 Prediction Error We tackle the classification problem using supervised learning methods due to the availability of true class labels of the available data. Hence it is possible to compute the correct and incorrect class predictions for the samples of data by comparing prediction results with the true labels and hence determine the misclassification error as the proportion of incorrectly classified samples. Conversely, the prediction accuracy is a measure of the proportion of samples that have been correctly classified by the models. 5.1.2 Hold-out and Cross-Validation Accuracy We determine the model performance using two types of validation methods. One being the hold-out accuracy and the other being cross-validation accuracy. The hold-out procedure consists of splitting the data into two sets namely the training and the testing set. The two sets of data are disjoint or mutually exclusive of each other and commonly the training set is the larger of the two sets. Hence, in this method the classifier is learned or trained on the training set without any exposure to the test set. The trained model is then applied to the test set and outputs predictions for each of the test sample classes which is compared with the true labels. Cross validation on the other hand involves splitting data into folds and training the classifier on data except a particular fold which is left for validation. This procedure is repeated across all folds and the final accuracy is representative of the average performance accuracy over all folds. We are interested to see how hold-out and cross validation accuracies of our chosen models compare against each other. [29]

Precision and recall are metrics which give us a picture about model performance and can be evaluated from the confusion matrix. Precision is a measure of how many of the predicted values in a class are correctly classified or actually part of the true labels of the class. Hence it is a measure of positive prediction by the model. Recall on the other hand is a measure of the amount of information correctly retrieved or in other words the number of samples correctly predicted. Models can be optimized based on a measure

which combines or balances both precision and recall. This is called F-measure which is a weighted average of the precision and recall of the model. Precision and recall can be computed from the confusion matrix to determine model performance. In a binary sense, the confusion matrix portrays the number of true positives, true negatives, false positives and false negatives. In the binary case, the confusion matrix is given by T P F N F P T N The precision and recall of the classifier can be computed from these values by the following formulations. Precision = TP / (TP+FP ) Recall = TP / (TP+FN ) This concept can be extended to the multiclass case for the precision and recall formulations.If M represents a confusion matrix for multiple classes, M being a k x k matrix where k is the number of classes Precision = Mii / P j Mji Recall = Mii / P i Mij