APPLIANCES ENERGY PREDICTION & BANK TERM LOAN DEPOSIT PREDICTION USING SUPPORT VECTOR MACHINES AND DECISION TREES

SUDHARSANA RAJASEKARAN

APPLIED MACHINE LEARNING

Assignment 2

Table of Contents

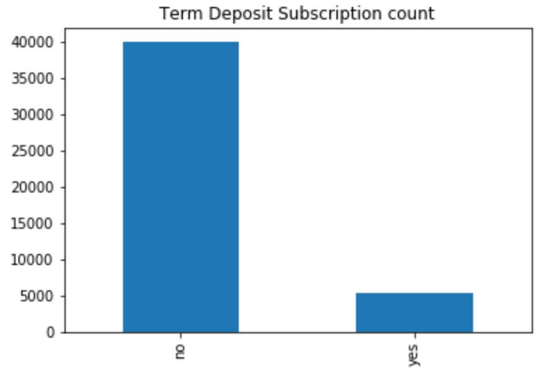
Introduction:

The data is related with direct marketing campaigns of a Portuguese banking institution and the dataset was derived from the UCI Machine learning repository. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

Exploratory Data Analysis:

Initial dataset analysis yielded the following results,

1. Number of Transactions: The dataset contains 45211 observation and 17 variables.
2. Missing Values: Fortunately, there are no missing values in the dataset.
3. Distribution of Classes: It can be observed that the distribution of classes is imbalanced. Therefore, the dataset must be balanced before running any algorithm.



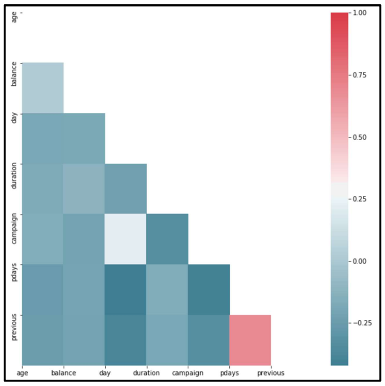


Figure 1. Term Deposit Subscription Count

Figure 2. Correlation Plot

4. Correlation Plot: The correlation plot tells us how far the dependent variables are correlated. It can be observed from the correlation plot that campaign and duration are highly negatively correlated with one another. The campaign and number of days since a client was last contacted are negatively correlated with one another.
5. Dataset Imbalance: The distribution of negative classes is high causing an imbalance in the distributions of classes. Hence to avoid this sample of the majority class is taken to make both the classes equally balanced.

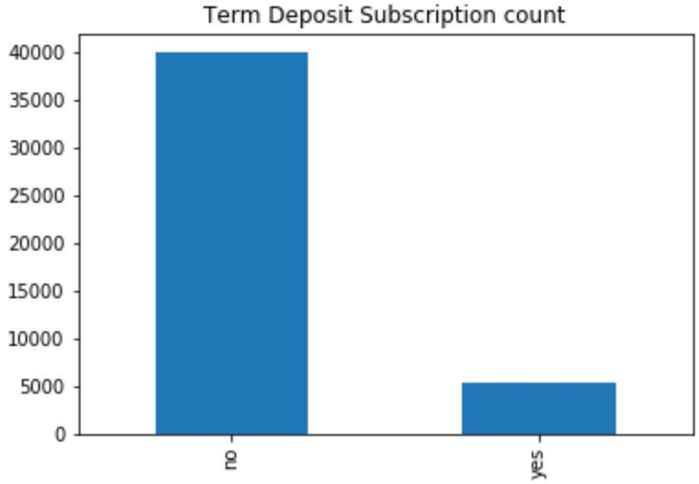


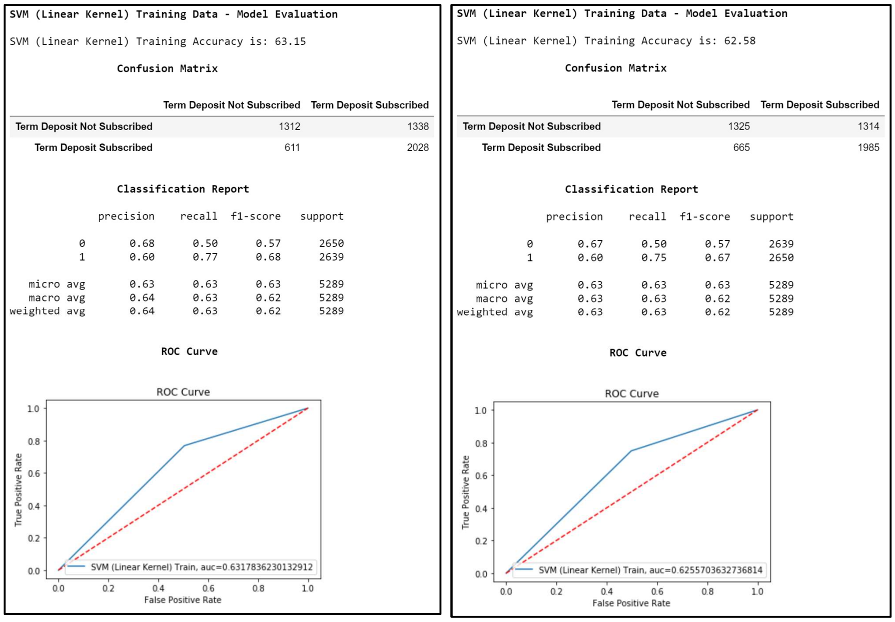
Figure 3. Distribution of Term deposit subscription

Term Deposit Subscription

SUPPORT VECTOR MACHINES

Linear Kernel

SVM with linear model has an accuracy of 63%. Although linear kernel can identify customers who have subscribed for term deposit with good accuracy, it does poorly in identifying customers who have not. It is important identifying these customers who have not subscribed any misclassification to that class is missed business opportunity for term deposit.



SVM (Linear Kernel) Training Data - Model Evaluation

SVM (Linear Kernel) Training Accuracy is: 63.15

Confusion Matrix

	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	1312	1338
Term Deposit Subscribed	611	2028

Classification Report

```
              precision    recall  f1-score   support

           0       0.68      0.50      0.57      2650
           1       0.60      0.77      0.68      2639

   micro avg       0.63      0.63      0.63      5289
   macro avg       0.64      0.63      0.62      5289
weighted avg       0.64      0.63      0.62      5289
```

ROC Curve

SVM (Linear Kernel) Training Data - Model Evaluation

SVM (Linear Kernel) Training Accuracy is: 62.58

Confusion Matrix

	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	1325	1314
Term Deposit Subscribed	665	1985

Classification Report

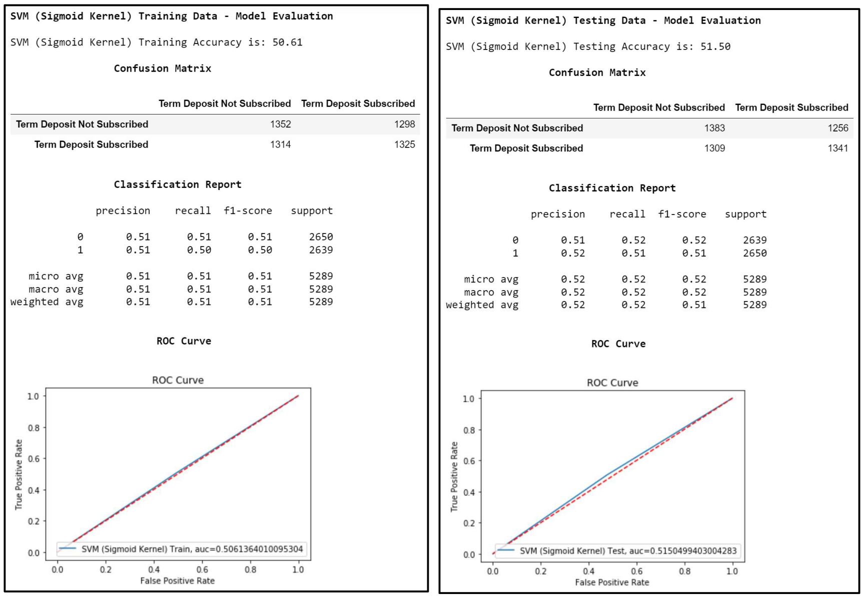
```
              precision    recall  f1-score   support

           0       0.67      0.50      0.57      2639
           1       0.60      0.75      0.67      2650

   micro avg       0.63      0.63      0.63      5289
   macro avg       0.63      0.63      0.62      5289
weighted avg       0.63      0.63      0.62      5289
```

ROC Curve

Figure 3. SVM Training and Test – Model Evaluation (Linear Kernel)

Sigmoid Kernel

Sigmoid kernel performs poorly over linear kernel. It can be observed from the ROC curve that the roc line of the sigmoid is almost the same as the default line, thus telling the model is not good for identifying the term deposit subscription.



SVM (Sigmoid Kernel) Training Data - Model Evaluation

SVM (Sigmoid Kernel) Training Accuracy is: 50.61

Confusion Matrix

	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	1352	1298
Term Deposit Subscribed	1314	1325

Classification Report

```
              precision    recall  f1-score   support

           0       0.51      0.51      0.51      2650
           1       0.51      0.50      0.50      2639

   micro avg       0.51      0.51      0.51      5289
   macro avg       0.51      0.51      0.51      5289
weighted avg       0.51      0.51      0.51      5289
```

ROC Curve

SVM (Sigmoid Kernel) Testing Data - Model Evaluation

SVM (Sigmoid Kernel) Testing Accuracy is: 51.50

Confusion Matrix

	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	1383	1256
Term Deposit Subscribed	1309	1341

Classification Report

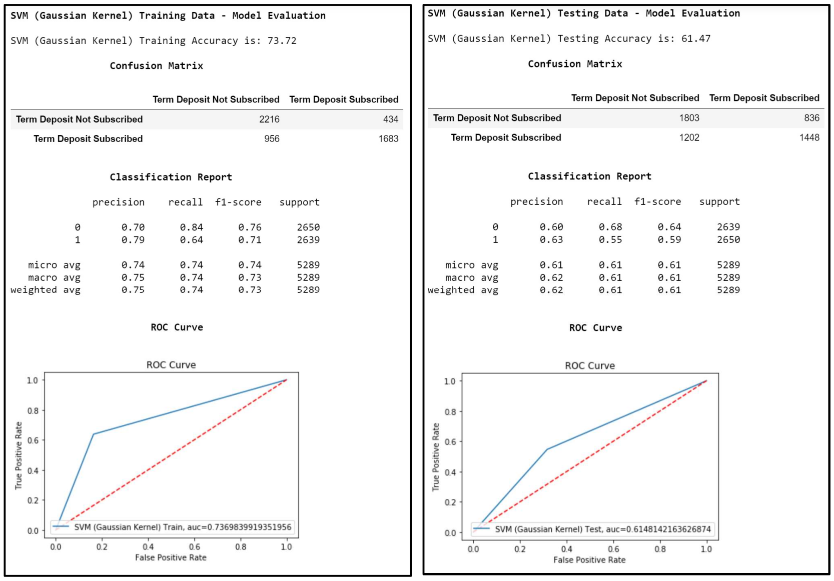
```
              precision    recall  f1-score   support

           0       0.51      0.52      0.52      2639
           1       0.52      0.51      0.51      2650

   micro avg       0.52      0.52      0.52      5289
   macro avg       0.52      0.52      0.52      5289
weighted avg       0.52      0.52      0.51      5289
```

ROC Curve

Figure 4. SVM Training and Test – Model Evaluation (Sigmoid Kernel)

Gaussian Kernel

Gaussian kernel of SVM overcomes the drawback of linear kernel by showing an improved performance of identifying the subscribers who have not subscribed for term deposit. Thus, adding to an overall training accuracy of 73%.



```
SVM (Gaussian Kernel) Training Data - Model Evaluation

SVM (Gaussian Kernel) Training Accuracy is: 73.72

                Confusion Matrix
```

	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	2216	434
Term Deposit Subscribed	956	1683

```
                Classification Report

             precision    recall  f1-score   support

           0       0.70      0.84      0.76      2650
           1       0.79      0.64      0.71      2639

   micro avg       0.74      0.74      0.74      5289
   macro avg       0.75      0.74      0.73      5289
weighted avg       0.75      0.74      0.73      5289

                      ROC Curve
```

```
SVM (Gaussian Kernel) Testing Data - Model Evaluation

SVM (Gaussian Kernel) Testing Accuracy is: 61.47

                Confusion Matrix
```

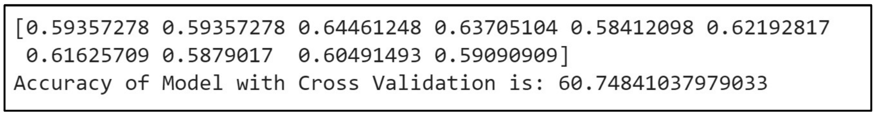
	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	1803	836
Term Deposit Subscribed	1202	1448

```
                Classification Report

             precision    recall  f1-score   support

           0       0.60      0.68      0.64      2639
           1       0.63      0.55      0.59      2650

   micro avg       0.61      0.61      0.61      5289
   macro avg       0.62      0.61      0.61      5289
weighted avg       0.62      0.61      0.61      5289

                      ROC Curve
```

Figure 5. SVM Training and Test – Model Evaluation (Gaussian Kernel)

K-fold Cross validated SVM

10-fold cross validated SVM model was built using gaussian kernel and the final mean accuracy of the model was found to be 60.74%.



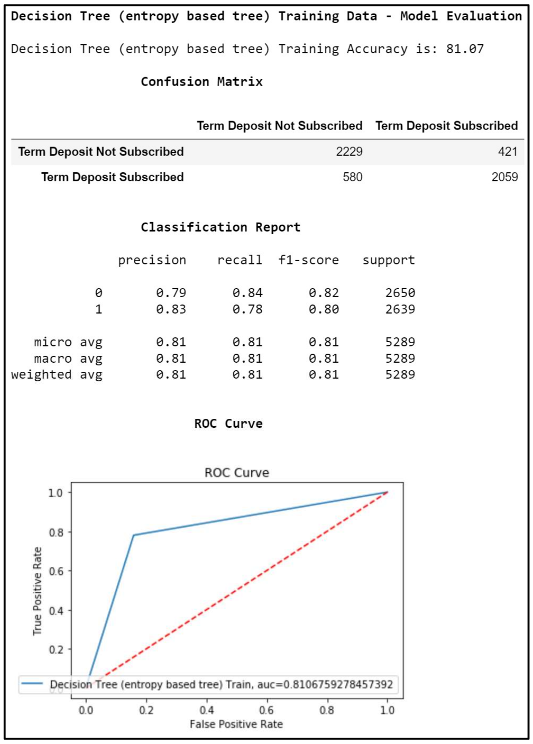
```
[0.59357278 0.59357278 0.64461248 0.63705104 0.58412098 0.62192817
 0.61625709 0.5879017  0.60491493 0.59090909]
Accuracy of Model with Cross Validation is: 60.74841037979033
```

Figure 7. SVM Training and Test – Model Evaluation (K-Fold)

Decision Trees:

Entropy based Decision Tree

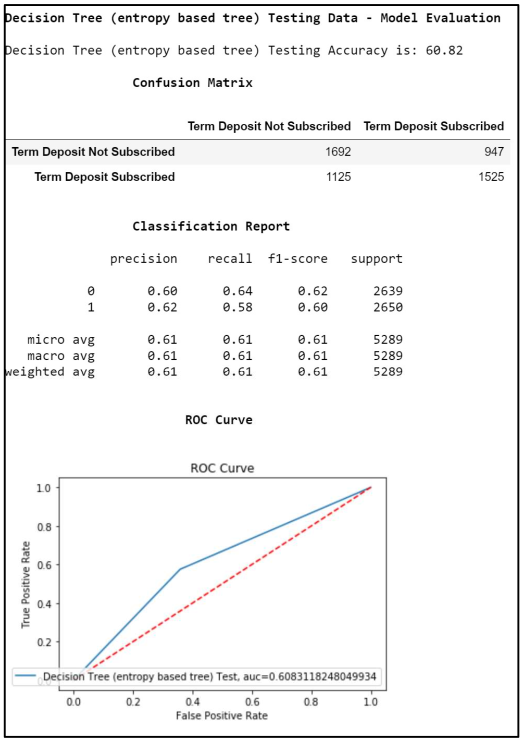
A decision tree without any pruning constraint was constructed for the term deposit classification. It can be observed that the tree grows without any constraint overfitting the training dataset. Thus, it poorly performs on the testing dataset. The model of decision tree has a training accuracy of 81% and testing accuracy of 64%.



```
Decision Tree (entropy based tree) Training Data - Model Evaluation

Decision Tree (entropy based tree) Training Accuracy is: 81.07

                 Confusion Matrix
```

	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	2229	421
Term Deposit Subscribed	580	2059

```
                 Classification Report

              precision    recall  f1-score   support

           0       0.79      0.84      0.82      2650
           1       0.83      0.78      0.80      2639

   micro avg       0.81      0.81      0.81      5289
   macro avg       0.81      0.81      0.81      5289
weighted avg       0.81      0.81      0.81      5289

                        ROC Curve
```



```
Decision Tree (entropy based tree) Testing Data - Model Evaluation

Decision Tree (entropy based tree) Testing Accuracy is: 60.82

                 Confusion Matrix
```

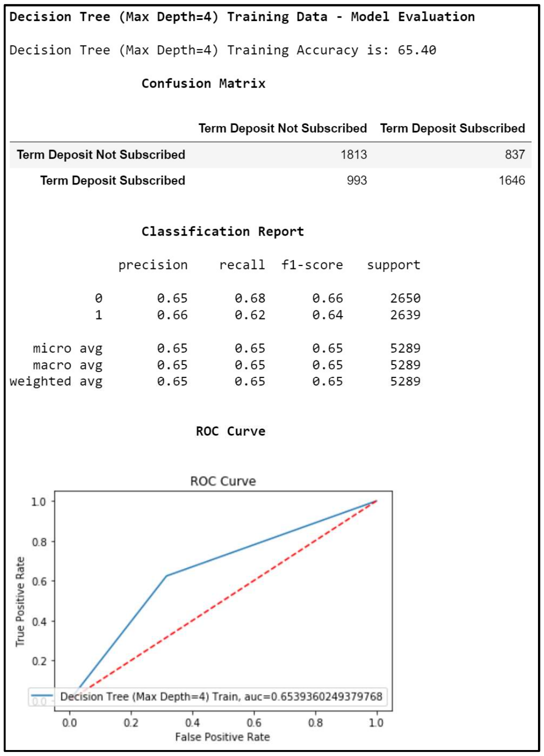
	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	1692	947
Term Deposit Subscribed	1125	1525

```
                 Classification Report

              precision    recall  f1-score   support

           0       0.60      0.64      0.62      2639
           1       0.62      0.58      0.60      2650

   micro avg       0.61      0.61      0.61      5289
   macro avg       0.61      0.61      0.61      5289
weighted avg       0.61      0.61      0.61      5289

                        ROC Curve
```

Figure 8. Decision Tree Training and Test – Model Evaluation (Entropy based)

Pruned Decision Tree

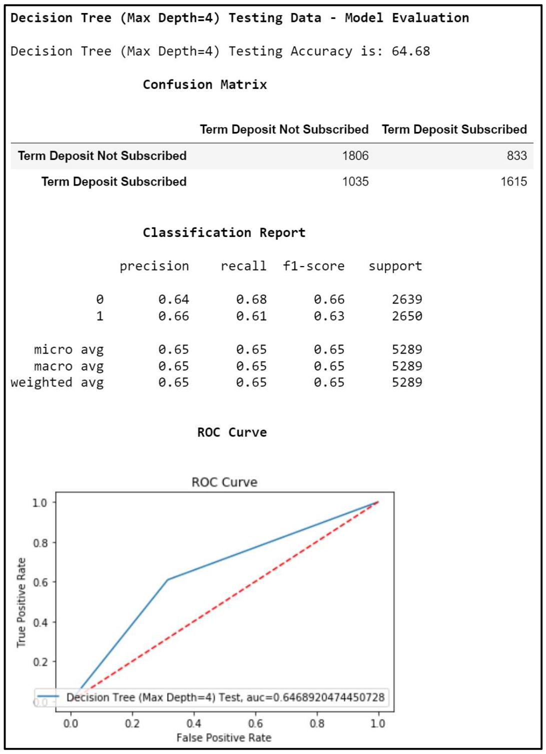
Pruning are important part of decision tree modelling. Pruning avoids overfitting of the model. It can be observed that the training and testing accuracies are almost the same. This tells us that the model is able to generalize well on the testing data set.



```
Decision Tree (Max Depth=4) Training Data - Model Evaluation

Decision Tree (Max Depth=4) Training Accuracy is: 65.40

                 Confusion Matrix
```

	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	1813	837
Term Deposit Subscribed	993	1646

```
                 Classification Report

              precision    recall  f1-score   support

           0       0.65      0.68      0.66      2650
           1       0.66      0.62      0.64      2639

   micro avg       0.65      0.65      0.65      5289
   macro avg       0.65      0.65      0.65      5289
weighted avg       0.65      0.65      0.65      5289

                        ROC Curve
```



```
Decision Tree (Max Depth=4) Testing Data - Model Evaluation

Decision Tree (Max Depth=4) Testing Accuracy is: 64.68

                 Confusion Matrix
```

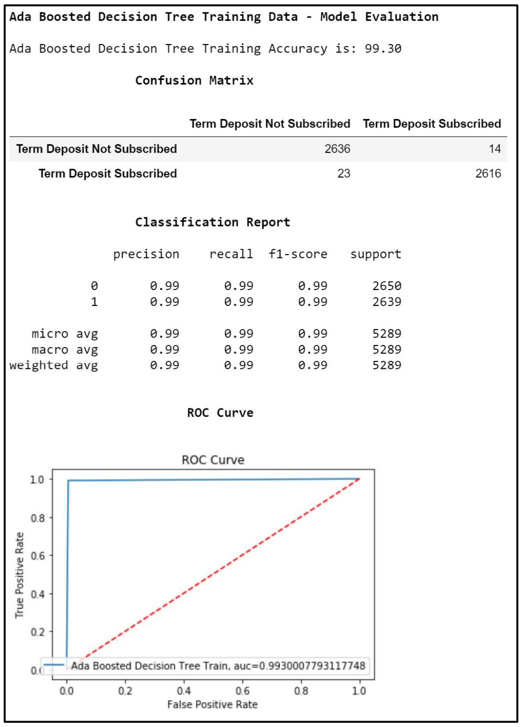
	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	1806	833
Term Deposit Subscribed	1035	1615

```
                 Classification Report

              precision    recall  f1-score   support

           0       0.64      0.68      0.66      2639
           1       0.66      0.61      0.63      2650

   micro avg       0.65      0.65      0.65      5289
   macro avg       0.65      0.65      0.65      5289
weighted avg       0.65      0.65      0.65      5289

                        ROC Curve
```

Figure 9. Decision Tree Training and Test – Model Evaluation (Pruned)

Ada Boosted Decision Tree

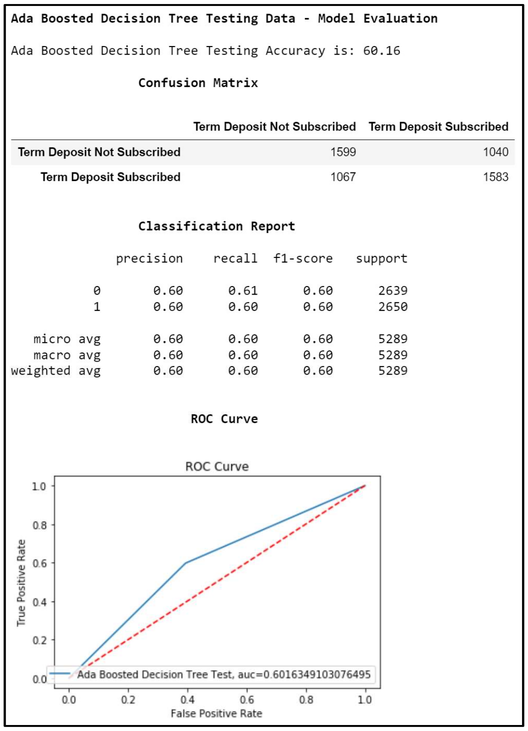
Boosting improves the performance of decision trees. The data boosted decision tree has an overall training accuracy of 99%. But however, it does poorly on the test dataset.



```
Ada Boosted Decision Tree Training Data - Model Evaluation

Ada Boosted Decision Tree Training Accuracy is: 99.30

                 Confusion Matrix
```

	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	2636	14
Term Deposit Subscribed	23	2616

```
                 Classification Report

              precision    recall  f1-score   support

           0       0.99      0.99      0.99      2650
           1       0.99      0.99      0.99      2639

   micro avg       0.99      0.99      0.99      5289
   macro avg       0.99      0.99      0.99      5289
weighted avg       0.99      0.99      0.99      5289

                      ROC Curve
```



```
Ada Boosted Decision Tree Testing Data - Model Evaluation

Ada Boosted Decision Tree Testing Accuracy is: 60.16

                 Confusion Matrix
```

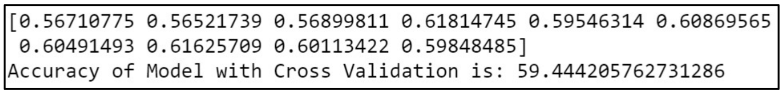
	Term Deposit Not Subscribed	Term Deposit Subscribed
Term Deposit Not Subscribed	1599	1040
Term Deposit Subscribed	1067	1583

```
                 Classification Report

              precision    recall  f1-score   support

           0       0.60      0.61      0.60      2639
           1       0.60      0.60      0.60      2650

   micro avg       0.60      0.60      0.60      5289
   macro avg       0.60      0.60      0.60      5289
weighted avg       0.60      0.60      0.60      5289

                      ROC Curve
```

Figure 10. Decision Tree Training and Test – Model Evaluation (Ada Boosted)

Cross Validated Decision Trees:

10-fold cross validated decision scored an overall mean accuracy of 59%.



```
[0.56710775 0.56521739 0.56899811 0.61814745 0.59546314 0.60869565
 0.60491493 0.61625709 0.60113422 0.59848485]
Accuracy of Model with Cross Validation is: 59.444205762731286
```

Figure 11. Decision Tree Training and Test – Model Evaluation (K-Fold)

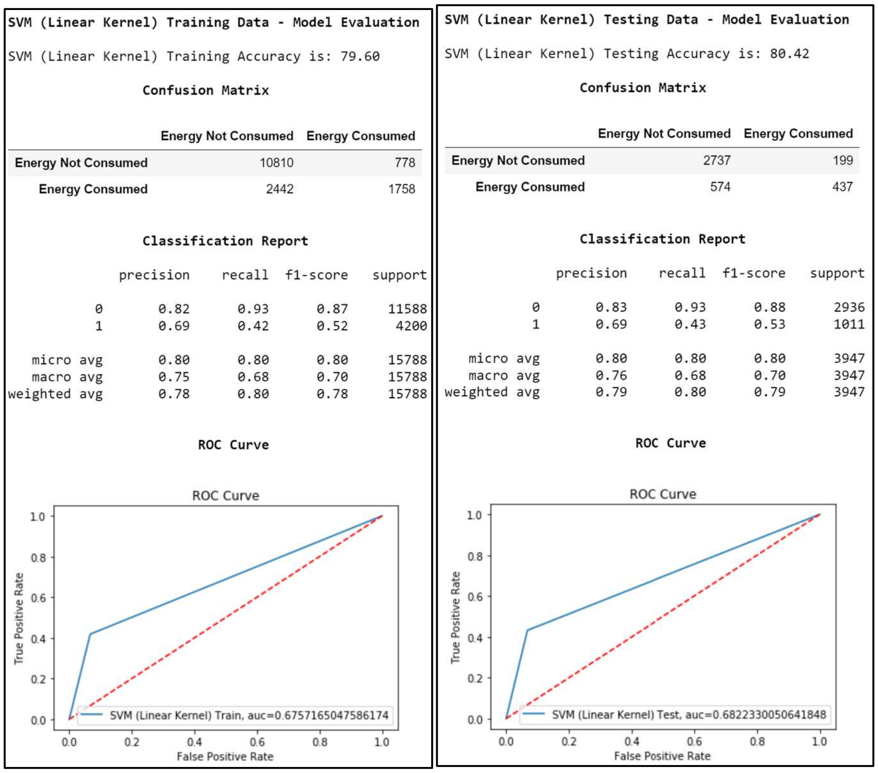
Appliances Energy Prediction

The Appliances Energy prediction dataset was derived from the UCI Machine learning repository. The dataset consists of house temperature and humidity conditions that were monitored with a ZigBee wireless sensor network. Each wireless node transmitted the temperature and humidity conditions around 3.3 min. Then, the wireless data was averaged for 10 minutes periods, which then becomes each of the rows in the dataset. The aim of the project is predicting the energy usage of the appliances using Linear and Logistic Regression by hyperparameter tuning of parameters

SUPPORT VECTOR MACHINES

Linear Kernel

The most basic kernel is the linear kernel. With linear as the kernel there is a training accuracy of 79% and a test error of 80.42%. From the confusion matrix, it can be observed that the misclassification rate for "Energy consumed" is high and hence could be mitigated by using different kernels.



```
SVM (Linear Kernel) Training Data - Model Evaluation

SVM (Linear Kernel) Training Accuracy is: 79.60

                Confusion Matrix
```

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	10810	778
Energy Consumed	2442	1758

```
              Classification Report

             precision    recall  f1-score   support

           0       0.82      0.93      0.87     11588
           1       0.69      0.42      0.52      4200

   micro avg       0.80      0.80      0.80     15788
   macro avg       0.75      0.68      0.70     15788
weighted avg       0.78      0.80      0.78     15788

                    ROC Curve
```

```
SVM (Linear Kernel) Testing Data - Model Evaluation

SVM (Linear Kernel) Testing Accuracy is: 80.42

                Confusion Matrix
```

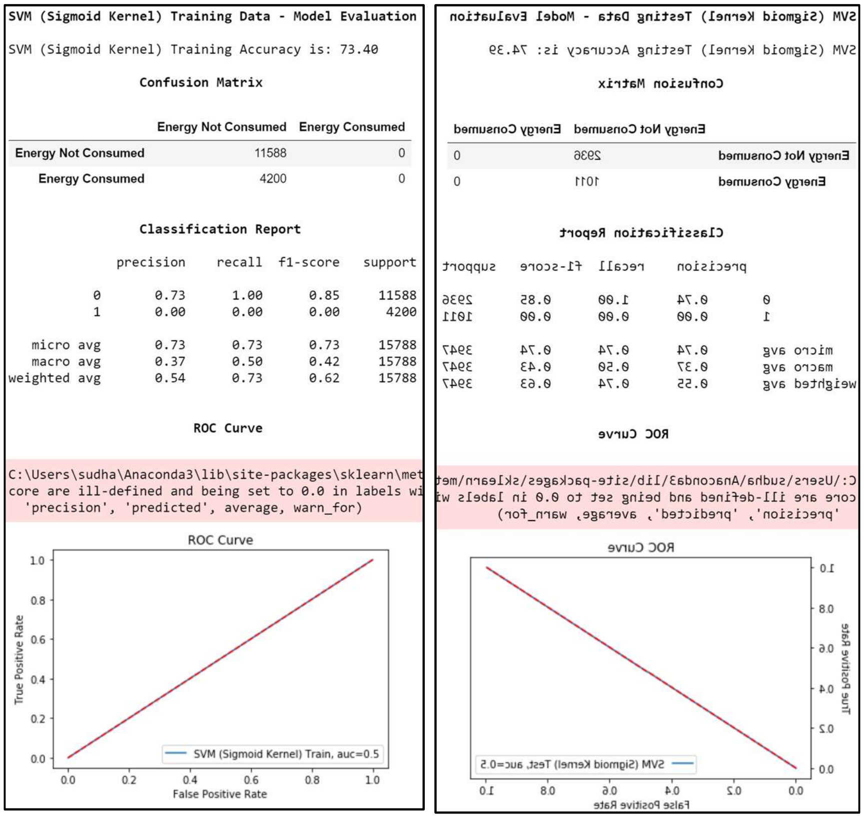
	Energy Not Consumed	Energy Consumed
Energy Not Consumed	2737	199
Energy Consumed	574	437

```
              Classification Report

             precision    recall  f1-score   support

           0       0.83      0.93      0.88      2936
           1       0.69      0.43      0.53      1011

   micro avg       0.80      0.80      0.80      3947
   macro avg       0.76      0.68      0.70      3947
weighted avg       0.79      0.80      0.79      3947

                    ROC Curve
```

Figure 12. SVM Training and Test – Model Evaluation (Linear Kernel)

Sigmoid Kernel

The sigmoid kernel shows and improvement over linear in identifying "Energy not consumed" class. But on the other hand, it is totally bad no identifying the "Energy consumed" classes. This model's weakness for identifying the "energy consumed" classes can be handled by using Gaussian kernel.



```
SVM (Sigmoid Kernel) Training Data - Model Evaluation

SVM (Sigmoid Kernel) Training Accuracy is: 73.40

                Confusion Matrix
```

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	11588	0
Energy Consumed	4200	0

```
              Classification Report

             precision    recall  f1-score   support

           0       0.73      1.00      0.85     11588
           1       0.00      0.00      0.00      4200

   micro avg       0.73      0.73      0.73     15788
   macro avg       0.37      0.50      0.42     15788
weighted avg       0.54      0.73      0.62     15788

                    ROC Curve

C:\Users\sudha\Anaconda3\lib\site-packages\sklearn\met
core are ill-defined and being set to 0.0 in labels wi
  'precision', 'predicted', average, warn_for)
```

```
SVM (Sigmoid Kernel) Testing Data - Model Evaluation

SVM (Sigmoid Kernel) Testing Accuracy is: 74.39

                Confusion Matrix
```

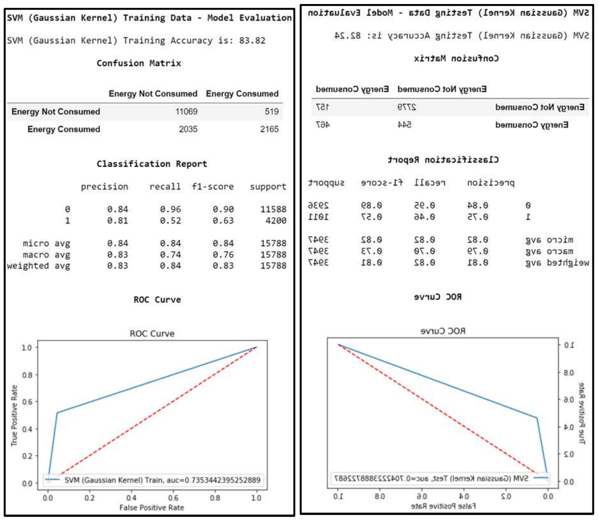
	Energy Not Consumed	Energy Consumed
Energy Not Consumed	2936	0
Energy Consumed	1011	0

```
              Classification Report

             precision    recall  f1-score   support

           0       0.74      1.00      0.85      2936
           1       0.00      0.00      0.00      1011

   micro avg       0.74      0.74      0.74      3947
   macro avg       0.37      0.50      0.43      3947
weighted avg       0.55      0.74      0.63      3947

                    ROC Curve

C:\Users\sudha\Anaconda3\lib\site-packages\sklearn\met
core are ill-defined and being set to 0.0 in labels wi
  'precision', 'predicted', average, warn_for)
```

Figure 13. SVM Training and Test – Model Evaluation (Sigmoid Kernel)

Gaussian Kernel

Gaussian kernel shows an improvement in identifying both linear and sigmoid kernels. Hence gaussian model here is good for energy consumption classification.



```
SVM (Gaussian Kernel) Training Data - Model Evaluation

SVM (Gaussian Kernel) Training Accuracy is: 83.82
```

Confusion Matrix

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	11069	519
Energy Consumed	2035	2165

```
Classification Report

              precision    recall  f1-score   support

           0       0.84      0.96      0.90     11588
           1       0.81      0.52      0.63      4200

   micro avg       0.84      0.84      0.84     15788
   macro avg       0.83      0.74      0.76     15788
weighted avg       0.83      0.84      0.83     15788

ROC Curve
```

```
SVM (Gaussian Kernel) Testing Data - Model Evaluation

SVM (Gaussian Kernel) Testing Accuracy is: 82.24
```

Confusion Matrix

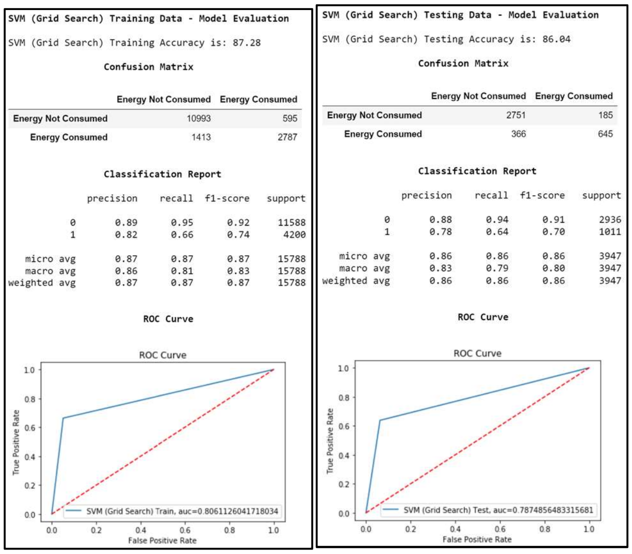
	Energy Not Consumed	Energy Consumed
Energy Not Consumed	2779	157
Energy Consumed	544	467

```
Classification Report

              precision    recall  f1-score   support

           0       0.84      0.95      0.89      2936
           1       0.75      0.46      0.57      1011

   micro avg       0.82      0.82      0.82      3947
   macro avg       0.79      0.70      0.73      3947
weighted avg       0.81      0.82      0.81      3947

ROC Curve
```

Figure 14. SVM Training and Test – Model Evaluation (Gaussian Kernel)

Grid Search SVM:

Using the Grid search method, the best parameters for this classification is found to be Gaussian kernel, C=1000, gamma=0.0001 . Grid search has helped us fine tune our classification model in improving the performance of our model. The model shows an improved performance after parameter tuning and showed an improved accuracy of 87% in training set and 86% in test set.



```
SVM (Grid Search) Training Data - Model Evaluation

SVM (Grid Search) Training Accuracy is: 87.28
```

Confusion Matrix

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	10993	595
Energy Consumed	1413	2787

```
Classification Report

              precision    recall  f1-score   support

           0       0.89      0.95      0.92     11588
           1       0.82      0.66      0.74      4200

   micro avg       0.87      0.87      0.87     15788
   macro avg       0.86      0.81      0.83     15788
weighted avg       0.87      0.87      0.87     15788

ROC Curve
```

```
SVM (Grid Search) Testing Data - Model Evaluation

SVM (Grid Search) Testing Accuracy is: 86.04
```

Confusion Matrix

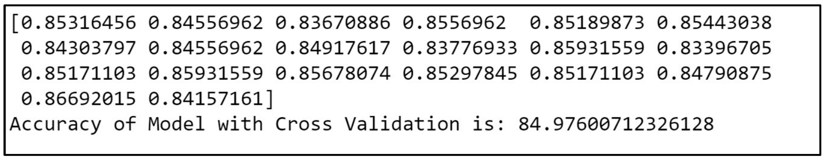
	Energy Not Consumed	Energy Consumed
Energy Not Consumed	2751	185
Energy Consumed	366	645

```
Classification Report

              precision    recall  f1-score   support

           0       0.88      0.94      0.91      2936
           1       0.78      0.64      0.70      1011

   micro avg       0.86      0.86      0.86      3947
   macro avg       0.83      0.79      0.80      3947
weighted avg       0.86      0.86      0.86      3947

ROC Curve
```

Figure 15. SVM Training and Test – Model Evaluation (Grid Search CV)

K-fold Cross validated SVM

With 20-fold cross validation we can achieve an accuracy of 84%



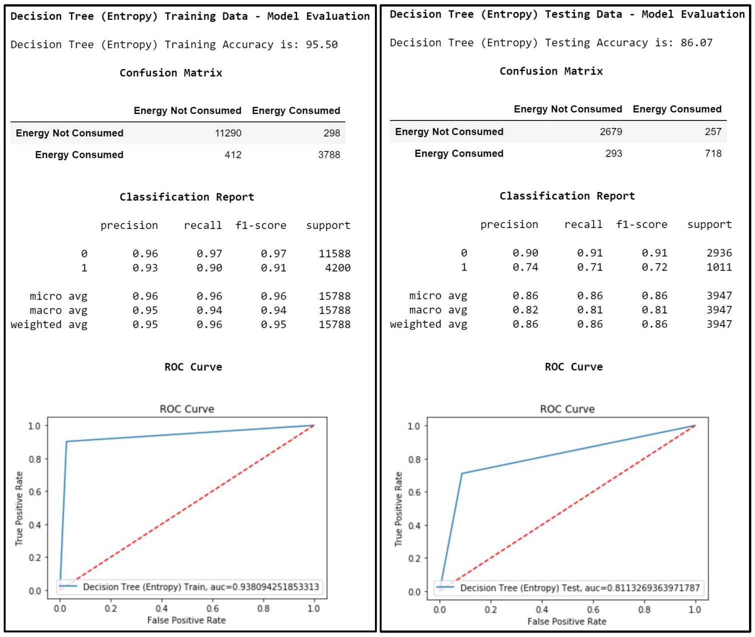
```
[0.85316456 0.84556962 0.83670886 0.8556962  0.85189873 0.85443038
 0.84303797 0.84556962 0.84917617 0.83776933 0.85931559 0.83396705
 0.85171103 0.85931559 0.85678074 0.85297845 0.85171103 0.84790875
 0.86692015 0.84157161]
Accuracy of Model with Cross Validation is: 84.97600712326128
```

Figure 16. SVM Training and Test – Model Evaluation (K-Fold)

Decision Trees:

Entropy based Decision Tree

Decision Tree Algorithm choose the highest Information gain to *split/construct* a Decision Tree. Here in this model the decision tree can grow relentlessly without any constraint. In general decision trees tend to over fit the test data and hence the decision tree that can grow without any condition perform poorly on the test data, since it will not be able to generalize well on the test data. Here in this model, even though the decision tree has an accuracy of 95% on the training data, it does not generalize well on the test data and thus produces a test accuracy of only 86%.



Decision Tree (Entropy) Training Data - Model Evaluation

Decision Tree (Entropy) Training Accuracy is: 95.50

Confusion Matrix

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	11290	298
Energy Consumed	412	3788

Classification Report

```
              precision    recall  f1-score   support

           0       0.96      0.97      0.97     11588
           1       0.93      0.90      0.91      4200

   micro avg       0.96      0.96      0.96     15788
   macro avg       0.95      0.94      0.94     15788
weighted avg       0.95      0.96      0.95     15788
```

ROC Curve

Decision Tree (Entropy) Testing Data - Model Evaluation

Decision Tree (Entropy) Testing Accuracy is: 86.07

Confusion Matrix

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	2679	257
Energy Consumed	293	718

Classification Report

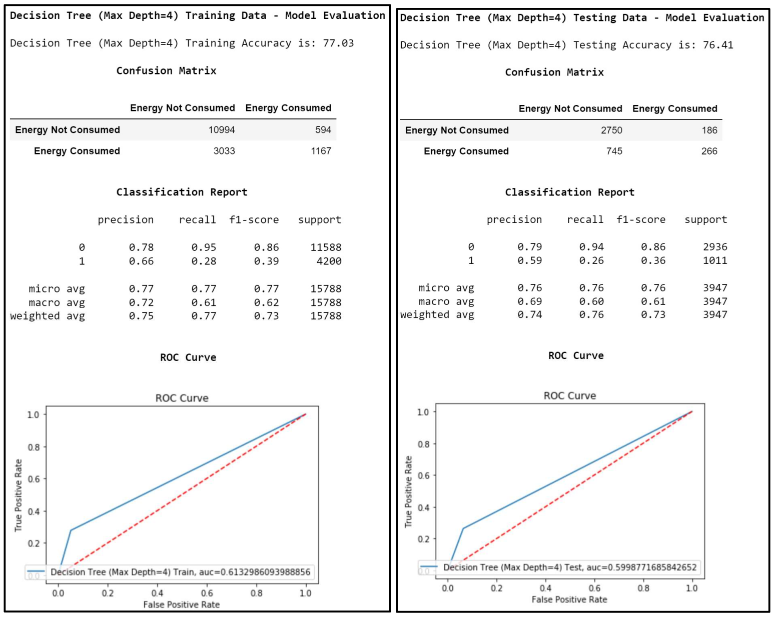
```
              precision    recall  f1-score   support

           0       0.90      0.91      0.91      2936
           1       0.74      0.71      0.72      1011

   micro avg       0.86      0.86      0.86      3947
   macro avg       0.82      0.81      0.81      3947
weighted avg       0.86      0.86      0.86      3947
```

ROC Curve

Figure 17. Decision Tree Training and Test – Model Evaluation (Entropy based)

Pruned Decision Tree

Pruning decision trees are very important as this could possibly avoid overfitting the training data. Here in the model the tree is pruned at a maximum depth of 4. Pruning makes the model generalize well on the training data. This additional flexibility comes with a price of training and testing accuracies being less that of an un-pruned tree. Here in the pruned decision tree, the training and testing accuracies are closer to each other, letting know the goodness of fit of the model.



Decision Tree (Max Depth=4) Training Data - Model Evaluation

Decision Tree (Max Depth=4) Training Accuracy is: 77.03

Confusion Matrix

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	10994	594
Energy Consumed	3033	1167

Classification Report

```
              precision    recall  f1-score   support

           0       0.78      0.95      0.86     11588
           1       0.66      0.28      0.39      4200

   micro avg       0.77      0.77      0.77     15788
   macro avg       0.72      0.61      0.62     15788
weighted avg       0.75      0.77      0.73     15788
```

ROC Curve

Decision Tree (Max Depth=4) Testing Data - Model Evaluation

Decision Tree (Max Depth=4) Testing Accuracy is: 76.41

Confusion Matrix

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	2750	186
Energy Consumed	745	266

Classification Report

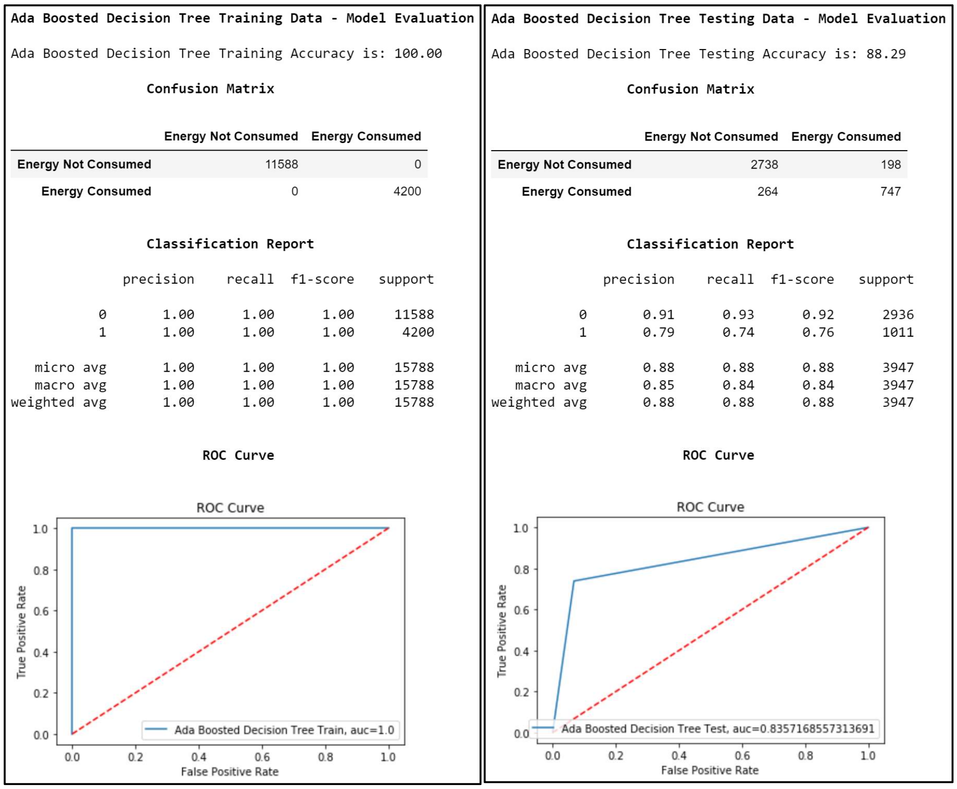
```
              precision    recall  f1-score   support

           0       0.79      0.94      0.86      2936
           1       0.59      0.26      0.36      1011

   micro avg       0.76      0.76      0.76      3947
   macro avg       0.69      0.60      0.61      3947
weighted avg       0.74      0.76      0.73      3947
```

ROC Curve

Figure 18. Decision Tree Training and Test – Model Evaluation (Pruned)

Ada Boosted Decision Tree

Boosting is another ensemble technique to create a collection of predictors. In this technique, learners are learned sequentially with early learners fitting simple models to the data and then analyzing data for errors. AdaBoost was the first successful boosting algorithm developed for binary classification. Ada boost has a training accuracy of 100%, that is the model is correctly able to identify all the observations in the training data set, it can also be seen that the test accuracy of boosted decision tree is 88% and ROC is 0.835 which is high when compared to decision tree with just pruning.



Ada Boosted Decision Tree Training Data - Model Evaluation

```
Ada Boosted Decision Tree Training Accuracy is: 100.00
```

Confusion Matrix

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	11588	0
Energy Consumed	0	4200

Classification Report

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     11588
           1       1.00      1.00      1.00      4200

   micro avg       1.00      1.00      1.00     15788
   macro avg       1.00      1.00      1.00     15788
weighted avg       1.00      1.00      1.00     15788
```

ROC Curve

Ada Boosted Decision Tree Testing Data - Model Evaluation

```
Ada Boosted Decision Tree Testing Accuracy is: 88.29
```

Confusion Matrix

	Energy Not Consumed	Energy Consumed
Energy Not Consumed	2738	198
Energy Consumed	264	747

Classification Report

```
              precision    recall  f1-score   support

           0       0.91      0.93      0.92      2936
           1       0.79      0.74      0.76      1011

   micro avg       0.88      0.88      0.88      3947
   macro avg       0.85      0.84      0.84      3947
weighted avg       0.88      0.88      0.88      3947
```

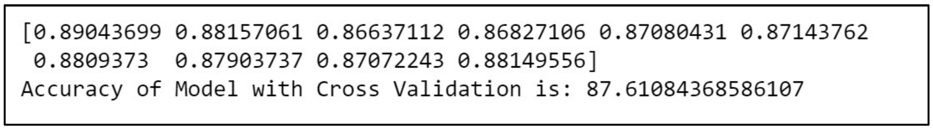
ROC Curve

Figure 19. Decision Tree Training and Test – Model Evaluation (Ada Boosted)

Cross Validated Decision Trees:

Cross-validation is a statistical technique for testing the performance of a Machine Learning model. A good cross validation method gives us a comprehensive measure of our model's performance throughout the whole dataset.

Here in the model, designed a 20-fold cross validation and the average training accuracy was found to be 87% which high when compared with a decision tree with pruning.



```
[0.89043699 0.88157061 0.86637112 0.86827106 0.87080431 0.87143762
 0.8809373  0.87903737 0.87072243 0.88149556]
Accuracy of Model with Cross Validation is: 87.61084368586107
```

Figure 20. Decision Tree Training and Test – Model Evaluation (K-Fold)