**Decision Tree**:

Building decision tree with selected features using c5.0 function:

Text

Description automatically generated with medium confidence

We know that the top 3 individual attributes that contribute to the decision tree model are Contract, Tenure, and Internet service from the summary output of the decision tree model. Features like OnlineSecurity, SeniorCitizen, PhoneService, PaperlessBilling, MultipleLines, gender, StreamingMovies, PaymentMethod and MonthlyCharges also play important roles in customer churn. We can generate a graphical representation of the decision tree for examining and viewing the underlying decision tree model results.

The following plot shows the results of the decision tree model. It shows that customers with loyal contracts (one or two years) are less likely to churn. The contract is the most significant feature to predict customer churn or not churn. When a client is on a month-to-month contract with DSL or No internet service, that customer is less likely to churn regardless of the length of the tenure. The customers who have fiber optic internet service with a higher tenure value (great than 14 months) are less likely to churn.

A picture containing bunch, long, different, several

Description automatically generated

**Training data performance:**

We can evaluate the decision tree model on training data. The output shows that the decision tree has 17 leaf nodes and resulted in 929 items being misclassified. The sensitivity is around 91.34% in the training data. The specificity is at 52.18% in the training data. The error rate for the model is 19.1% in the training data. We have an overall prediction accuracy of 80.9% which is relatively accurate.

A picture containing text, receipt

Description automatically generated

**Evaluating decision tree model performance (Validation data)**

We will also check prediction accuracy to see how well our decision tree model predicts churn in the validation subset. We performed the following confusion matrix to show the classification accuracy.

Text

Description automatically generated with medium confidence

The diagonal entries give us the correct predictions. The decision tree model performs better at predicting non-churning customers (1414 vs. 285 outcomes). The decision tree model below is fairly accurate in predicting the churning or non-churning customers. The sensitivity is about 91.23% and the specificity is around 50.89% in the test data. It has accuracy of around (1414 + 285)/2110 \*100% = 80.52% in the validation subset.

A picture containing text, receipt

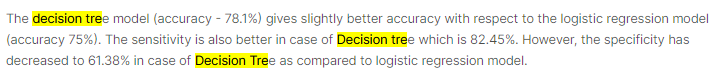
Description automatically generated

After analyzing the decision tree model, we can see that sensitivity (91.35%) and specificity (55.71%) in logistic regression model are both relatively larger than sensitivity (91.23%) and specificity (50.89%) in decision tree model. The prediction accuracy (81.90%) of the logistic regression slightly exceeds the decision tree model (80.52%). Apparently, the logistic regression model outperforms the decision tree model.

In the ROC plot below, we can see the dashed line in the diagonal to show the ROC curve of a random predictor. The random predictor is used to show whether a model is useful or not. Both curves are good since they lie above the dashed line. Logistic regression model is superior to decision tree model because at all cut-offs the true positive rate is slightly higher, and the false positive rate is slightly lower than for decision tree model. The AUC (area under the curve) for logistic regression model (85.6%) is slightly larger than the AUC for decision tree model (81.3%). The higher the AUC, the better the model is at correctly predicting churning or non-churning customers. The closer a ROC curve is to the upper left corner, the more efficient is the model. Hence, the logistic regression model performs better in churn analysis.

Chart, diagram

Description automatically generated



Finally , Logistic Regression with a cutoff probability value of 53.34% gives us better values of accuracy and specificity but lower in sensitivity.

First of all, decision Trees can easily use categorical features with entropy and gini index, but (Pearson) correlation is defined for only numeric attributes. Using correlation with one-hot encoded versions of these categorical features also makes no sense. Secondly, In classification tasks, correlation between a, let's say binary output (e.g. 0,10,1), and a numeric attribute may not make much sense as it normally does in regression problems.