Rahul Kumar

ZZSC5836 DATA MINING & MACHINE LEARNING ASSESSMENT 2: PROJECT I

Mining IBM Customer Churn Data

1. INTRODUCTION

The Report analyses the various accuracy of predictions done by both White Box and Black Box Models. Data Analysis and Hyperparameters are key for such models and their tuning can greatly affect the final result.

2. UNDERSTANDING THE DATA SET

The IBM Customer Churn Data is a decent dataset to analyse, and here are some key characteristics of it. The data set required no null value clean up. Most of the columns were object types. SeniorCitizen and tenure were integers. MonthlyCharges TotalCharges were floats.

Some interesting facts about the data are:

- The Churn Category was made up of 73.4% No values and 26.6% Yes Values.
- The average of Monthly Charges across the Dataset is 64.79, and it has a standard deviation of 30.09.
- Gender is evenly split with 50.5% being Male, and 49.5% being Female.
- Each Customer Entry is Unique.
- The longest tenure is 72 by a customer and shortest is 1.
- Male and Female customers had nearly equally long average tenures, 32.29 for females and 32.55 for males.

3. TASK-1: WHITE BOX MODELS

3.1 DATA PREPARATION

The data was well prepared prior to use so there was no data clean up required. Category Values were unique. Alongside this, the data had no null values, and as a result the entire dataset can be used for prediction. As the predicted column was churn, it was noticed that the data was heavily imbalanced. This was ignored for the scope of this project, however to improve prediction and possible alternative would be to use a resampling technique. Random Undersampling would be recommended to balance the dataset out. The only preparation that was carried out was ordinal encoding, where categorical data was replaced with integers.

3.2 METHODOLOGY

An important factor to consider before deciding on a model, would be to derive the feature columns of that will be the predictors of Churn. This was done after ordinal encoding, and then presented as a correlation heatmap. As shown in the Appendix, Figure 1. By looking at the Churn_Code row, we can easily determine feature columns that would have an effect on Churn.

- SeniorCitizen
- tenure
- MonthlyCharges
- TotalCharges
- Partner
- Dependents
- OnlineSecurity
- OnlineBackup
- DeviceProtection
- TechSupport
- Contract
- PaperlessBilling
- PaymentMethod

These were selected of having a correlation value greater than 0.1. The classifying model used was a Decision Tree Classifier as the problem scenario requires result to be split into classes or categories.

3.3 Building & Evaluating the Model

Primarily building the model was simple. The dataset was split use train_test_split, with the default split value of 75% to 25%. To discover the ideal hyperparameter for the model GridSearchCV was used, and a small range of numbers provided to reduce computational time. The hyperparameters tuned were criterion, min_samples_leaf, max_depth and splitter. After which, accuracy was determined through the use of a confusion matrix (Shown in the Appendix, Table 1.) The model's accuracy can be determined as follows

- Accuracy is 85.38%
- Misclassification Rate is 14.62%
- True Positive Rate is 67.54%
- False Positive Rate is 8.31%
- Precision is 74.16%
- Prevalence is 26.11%

Some interesting rules/nodes derived in this tree are pictured in the appendix and summarised below:

- If Monthly Charges ≤ 58.675, Class is No, with a leaf purity of 93.13%, Samples = 487
- If Paperless Billing ≤ 0.5, Class is No, with a leaf purity of 90.18%, Samples = 163
- If Total Charges ≤ 3246.375, Class is No, with a leaf purity of 99.70%, Samples = 335
- If Total Charges ≤ 1277.3, Class is Yes, with a leaf purity of 81.39%, Samples = 231
- If Tenure ≤ 3.5, Class is Yes, with a leaf purity of 83.18%, Samples = 898
- If Paperless Billing ≤ 0.5, Class is Yes, with a leaf purity of 96.83%, Samples = 126

The final rule tree can be seen in the Appendix.

3.4 Conclusion

The hyperparameters were key in getting an ideal model, and gave a model with an accuracy of 85%. However, a greater accuracy can be obtained with more tuning, where computational time is not a factor.

4. Task-2: Black Box Models

4.1 Data Preparation

Data Preparation was the same as outlined in 3.1.

4.2 Methodology

Feature columns were chosen in the same method as shown in 3.2. Random Forest Classifier was chosen as the model as its similarity to Decision Tree Classifier. The Hyperparameter tuning will be similar.

4.3 Building & Evaluating the Model

To tune the Hyperparameters, once again GridSearchCV was carried out, after the data was split using train_test_split. The Hyperparameters tuned were n_estimators, criterion, bootstrap, min_samples_leaf and max_depth. These parameters were slightly tuned to simple constraints to reduce computational time. Afterwhich to achieve the exact requirements of Churn value of Yes to have a 80% precision and 60%, there were some manual tuning done. The min_samples_leaf was increased, and the n_estimators was reduced to 200, from the base value given by GridSearchCV. Finally min_weight_fraction_leaf was tuned to achieve the desired result.

Evaluation was done using k-fold cross validation, with one left out. The X training data was split into 10 splits and shuffle being true. The final accuracy was judged to be adequate with average score across each fold being 85.38% and ROC-AUC score being 0.90. This model was then used to predict Churn based on the X testing data, and classification report was generated shown in the Appendix in Table 2. It can be seen that Churn Yes is at 80% precision and 63% recall.It should be noted that GridSearchCV or RandomSearchCV can also be used to tune the parameters to achieve better precision scores or Recall Scores, which can be computationally intensive, and is recommended if a specific model requirement is needed, over hand tuning.

4.4 Conclusion

The tuning of the hyperparameters were the most time consuming part of this model, and with the fact that a user is unable to look at the specifics of the Model, hence the term black box, makes it difficult to tune to specific scenarios. This is counteracted with GridSearchCV using parameters to tune to scoring. KFold Cross Validation also gives a sense of confidence to the model.

5.Conclusion

Both models provided a great understanding to the differences between Black Box and White Box models. Black box models are better with complex problems at the cost of interpretability and explainability. How you optimise the hyperparameters and evaluate the accuracy is a big part of the usefulness of the model.3

Rahul Kumar

Appendix

Task 1

FIGURE 1. TASK 1 HEAT CORRELATION MAP

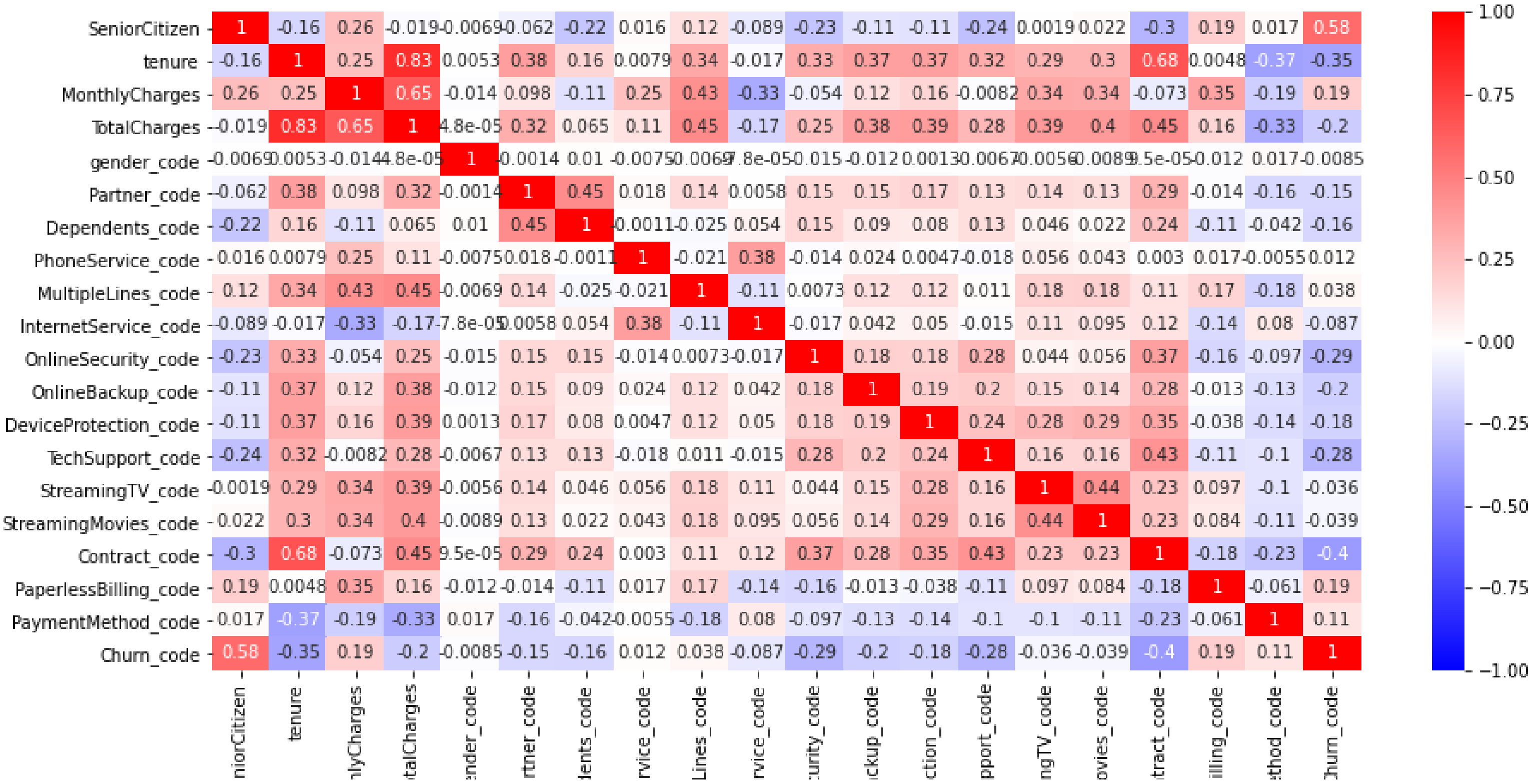

TABLE 1. TASK 1 CONFUSION MATRIX

	Predicted No	Predicted Yes	
Actual No	TN = 1340	FP = 108	1299
Actual Yes	FN = 149	TP = 310	459
	1340	418	

Task 2

TABLE 2. TASK 2- CLASSIFICATION REPORT

	Precision	Recall	F1	Support
No	0.88	0.94	0.91	1299
Yes	0.8	0.63	0.7	459
Accuracy			0.86	1758
Macro Avg	0.84	0.79	0.81	1758
Weighted Avg	0.86	0.86	0.86	1758

Rahul Kumar

TASK 1 Tree

```
digraph Tree {
node [shape=box] ;
0 [label="SeniorCitizen <= 0.5\nentropy = 0.838\nsamples = 5274\nvalue
= [3864, 1410]\nclass = No"] ;
1 [label="Contract_code <= 0.5\nentropy = 0.478\nsamples = 3767\nvalue
= [3379, 388]\nclass = No"] ;
0 -> 1 [labeldistance=2.5, labelangle=45, headlabel="True"] ;
2 [label="tenure <= 6.5\nentropy = 0.708\nsamples = 1709\nvalue =
[1379, 330]\nclass = No"] ;
1 -> 2 ;
3 [label="OnlineSecurity_code <= 0.5\nentropy = 0.885\nsamples =
601\nvalue = [419, 182]\nclass = No"] ;
2 -> 3 ;
4 [label="MonthlyCharges <= 66.75\nentropy = 0.987\nsamples =
340\nvalue = [193, 147]\nclass = No"] ;
3 -> 4 ;
5 [label="PaperlessBilling_code <= 0.5\nentropy = 0.855\nsamples =
161\nvalue = [116, 45]\nclass = No"] ;
4 -> 5 ;
6 [label="entropy = 0.597\nsamples = 69\nvalue = [59, 10]\nclass = No"]
;
5 -> 6 ;
7 [label="entropy = 0.958\nsamples = 92\nvalue = [57, 35]\nclass = No"]
;
5 -> 7 ;
8 [label="TotalCharges <= 84.45\nentropy = 0.986\nsamples = 179\nvalue
= [77, 102]\nclass = Yes"] ;
4 -> 8 ;
9 [label="entropy = 0.782\nsamples = 56\nvalue = [13, 43]\nclass =
Yes"] ;
8 -> 9 ;
10 [label="MonthlyCharges <= 80.85\nentropy = 0.999\nsamples =
123\nvalue = [64, 59]\nclass = No"] ;
8 -> 10 ;
11 [label="entropy = 0.928\nsamples = 64\nvalue = [42, 22]\nclass =
No"] ;
10 -> 11 ;
12 [label="entropy = 0.953\nsamples = 59\nvalue = [22, 37]\nclass =
Yes"] ;
10 -> 12 ;
13 [label="TotalCharges <= 36.325\nentropy = 0.569\nsamples =
261\nvalue = [226, 35]\nclass = No"] ;
3 -> 13 ;
14 [label="entropy = 0.71\nsamples = 98\nvalue = [79, 19]\nclass = No"]
;
13 -> 14 ;
15 [label="PaperlessBilling_code <= 0.5\nentropy = 0.463\nsamples =
163\nvalue = [147, 16]\nclass = No"] ;
13 -> 15 ;
16 [label="entropy = 0.34\nsamples = 95\nvalue = [89, 6]\nclass = No"]
;
15 -> 16 ;
```

```
17 [label="entropy = 0.602\nsamples = 68\nvalue = [58, 10]\nclass =
No"] ;
15 -> 17 ;
18 [label="MonthlyCharges <= 70.425\nentropy = 0.567\nsamples =
1108\nvalue = [960, 148]\nclass = No"] ;
2 -> 18 ;
19 [label="MonthlyCharges <= 19.975\nentropy = 0.377\nsamples =
547\nvalue = [507, 40]\nclass = No"] ;
18 -> 19 ;
20 [label="entropy = 0.0\nsamples = 60\nvalue = [60, 0]\nclass = No"] ;
19 -> 20 ;
21 [label="MonthlyCharges <= 58.675\nentropy = 0.41\nsamples =
487\nvalue = [447, 40]\nclass = No"] ;
19 -> 21 ;
22 [label="PaymentMethod_code <= 2.5\nentropy = 0.466\nsamples =
364\nvalue = [328, 36]\nclass = No"] ;
21 -> 22 ;
23 [label="Dependents_code <= 0.5\nentropy = 0.539\nsamples =
243\nvalue = [213, 30]\nclass = No"] ;
22 -> 23 ;
24 [label="TotalCharges <= 1056.35\nentropy = 0.618\nsamples =
163\nvalue = [138, 25]\nclass = No"] ;
23 -> 24 ;
25 [label="entropy = 0.713\nsamples = 92\nvalue = [74, 18]\nclass =
No"] ;
24 -> 25 ;
26 [label="entropy = 0.465\nsamples = 71\nvalue = [64, 7]\nclass = No"]
;
24 -> 26 ;
27 [label="entropy = 0.337\nsamples = 80\nvalue = [75, 5]\nclass = No"]
;
23 -> 27 ;
28 [label="tenure <= 17.5\nentropy = 0.285\nsamples = 121\nvalue =
[115, 6]\nclass = No"] ;
22 -> 28 ;
29 [label="entropy = 0.4\nsamples = 63\nvalue = [58, 5]\nclass = No"] ;
28 -> 29 ;
30 [label="entropy = 0.126\nsamples = 58\nvalue = [57, 1]\nclass = No"]
;
28 -> 30 ;
31 [label="TotalCharges <= 1423.225\nentropy = 0.207\nsamples =
123\nvalue = [119, 4]\nclass = No"] ;
21 -> 31 ;
32 [label="entropy = 0.313\nsamples = 71\nvalue = [67, 4]\nclass = No"]
;
31 -> 32 ;
33 [label="entropy = 0.0\nsamples = 52\nvalue = [52, 0]\nclass = No"] ;
31 -> 33 ;
34 [label="tenure <= 13.5\nentropy = 0.707\nsamples = 561\nvalue =
[453, 108]\nclass = No"] ;
18 -> 34 ;
35 [label="MonthlyCharges <= 80.025\nentropy = 0.946\nsamples =
118\nvalue = [75, 43]\nclass = No"] ;
34 -> 35 ;
36 [label="entropy = 0.804\nsamples = 57\nvalue = [43, 14]\nclass =
No"] ;
35 -> 36 ;
```

```
37 [label="entropy = 0.998\nsamples = 61\nvalue = [32, 29]\nclass =
No"] ;
35 -> 37 ;
38 [label="PaymentMethod_code <= 1.5\nentropy = 0.602\nsamples =
443\nvalue = [378, 65]\nclass = No"] ;
34 -> 38 ;
39 [label="TechSupport_code <= 1.0\nentropy = 0.378\nsamples =
191\nvalue = [177, 14]\nclass = No"] ;
38 -> 39 ;
40 [label="MonthlyCharges <= 91.0\nentropy = 0.449\nsamples =
128\nvalue = [116, 12]\nclass = No"] ;
39 -> 40 ;
41 [label="entropy = 0.303\nsamples = 74\nvalue = [70, 4]\nclass = No"]
;
40 -> 41 ;
42 [label="entropy = 0.605\nsamples = 54\nvalue = [46, 8]\nclass = No"]
;
40 -> 42 ;
43 [label="entropy = 0.203\nsamples = 63\nvalue = [61, 2]\nclass = No"]
;
39 -> 43 ;
44 [label="TotalCharges <= 4465.575\nentropy = 0.727\nsamples =
252\nvalue = [201, 51]\nclass = No"] ;
38 -> 44 ;
45 [label="MonthlyCharges <= 83.85\nentropy = 0.801\nsamples =
197\nvalue = [149, 48]\nclass = No"] ;
44 -> 45 ;
46 [label="entropy = 0.573\nsamples = 59\nvalue = [51, 8]\nclass = No"]
;
45 -> 46 ;
47 [label="tenure <= 29.5\nentropy = 0.869\nsamples = 138\nvalue = [98,
40]\nclass = No"] ;
45 -> 47 ;
48 [label="entropy = 0.918\nsamples = 84\nvalue = [56, 28]\nclass =
No"] ;
47 -> 48 ;
49 [label="entropy = 0.764\nsamples = 54\nvalue = [42, 12]\nclass =
No"] ;
47 -> 49 ;
50 [label="entropy = 0.305\nsamples = 55\nvalue = [52, 3]\nclass = No"]
;
44 -> 50 ;
51 [label="MonthlyCharges <= 49.3\nentropy = 0.185\nsamples =
2058\nvalue = [2000, 58]\nclass = No"] ;
1 -> 51 ;
52 [label="PaymentMethod_code <= 0.5\nentropy = 0.025\nsamples =
816\nvalue = [814, 2]\nclass = No"] ;
51 -> 52 ;
53 [label="TotalCharges <= 1059.25\nentropy = 0.077\nsamples =
211\nvalue = [209, 2]\nclass = No"] ;
52 -> 53 ;
54 [label="entropy = 0.149\nsamples = 94\nvalue = [92, 2]\nclass = No"]
;
53 -> 54 ;
55 [label="entropy = 0.0\nsamples = 117\nvalue = [117, 0]\nclass = No"]
;
53 -> 55 ;
```

```
56 [label="entropy = 0.0\nsamples = 605\nvalue = [605, 0]\nclass = No"]
;
52 -> 56 ;
57 [label="Contract_code <= 1.5\nentropy = 0.265\nsamples = 1242\nvalue
= [1186, 56]\nclass = No"] ;
51 -> 57 ;
58 [label="MonthlyCharges <= 101.95\nentropy = 0.368\nsamples =
582\nvalue = [541, 41]\nclass = No"] ;
57 -> 58 ;
59 [label="TotalCharges <= 3711.3\nentropy = 0.277\nsamples =
481\nvalue = [458, 23]\nclass = No"] ;
58 -> 59 ;
60 [label="TotalCharges <= 2316.175\nentropy = 0.362\nsamples =
304\nvalue = [283, 21]\nclass = No"] ;
59 -> 60 ;
61 [label="tenure <= 22.5\nentropy = 0.215\nsamples = 146\nvalue =
[141, 5]\nclass = No"] ;
60 -> 61 ;
62 [label="entropy = 0.396\nsamples = 64\nvalue = [59, 5]\nclass = No"]
;
61 -> 62 ;
63 [label="entropy = 0.0\nsamples = 82\nvalue = [82, 0]\nclass = No"] ;
61 -> 63 ;
64 [label="TotalCharges <= 3185.95\nentropy = 0.473\nsamples =
158\nvalue = [142, 16]\nclass = No"] ;
60 -> 64 ;
65 [label="OnlineSecurity_code <= 1.0\nentropy = 0.573\nsamples =
103\nvalue = [89, 14]\nclass = No"] ;
64 -> 65 ;
66 [label="entropy = 0.722\nsamples = 50\nvalue = [40, 10]\nclass =
No"] ;
65 -> 66 ;
67 [label="entropy = 0.386\nsamples = 53\nvalue = [49, 4]\nclass = No"]
;
65 -> 67 ;
68 [label="entropy = 0.225\nsamples = 55\nvalue = [53, 2]\nclass = No"]
;
64 -> 68 ;
69 [label="PaperlessBilling_code <= 0.5\nentropy = 0.089\nsamples =
177\nvalue = [175, 2]\nclass = No"] ;
59 -> 69 ;
70 [label="entropy = 0.191\nsamples = 68\nvalue = [66, 2]\nclass = No"]
;
69 -> 70 ;
71 [label="entropy = 0.0\nsamples = 109\nvalue = [109, 0]\nclass = No"]
;
69 -> 71 ;
72 [label="MonthlyCharges <= 106.725\nentropy = 0.676\nsamples =
101\nvalue = [83, 18]\nclass = No"] ;
58 -> 72 ;
73 [label="entropy = 0.76\nsamples = 50\nvalue = [39, 11]\nclass = No"]
;
72 -> 73 ;
74 [label="entropy = 0.577\nsamples = 51\nvalue = [44, 7]\nclass = No"]
;
72 -> 74 ;
75 [label="OnlineSecurity_code <= 1.0\nentropy = 0.156\nsamples =
660\nvalue = [645, 15]\nclass = No"] ;
```

```
57 -> 75 ;
76 [label="OnlineBackup_code <= 1.0\nentropy = 0.295\nsamples =
192\nvalue = [182, 10]\nclass = No"] ;
75 -> 76 ;
77 [label="entropy = 0.439\nsamples = 66\nvalue = [60, 6]\nclass = No"]
;
76 -> 77 ;
78 [label="tenure <= 68.5\nentropy = 0.203\nsamples = 126\nvalue =
[122, 4]\nclass = No"] ;
76 -> 78 ;
79 [label="entropy = 0.103\nsamples = 74\nvalue = [73, 1]\nclass = No"]
;
78 -> 79 ;
80 [label="entropy = 0.318\nsamples = 52\nvalue = [49, 3]\nclass = No"]
;
78 -> 80 ;
81 [label="TotalCharges <= 6573.475\nentropy = 0.085\nsamples =
468\nvalue = [463, 5]\nclass = No"] ;
75 -> 81 ;
82 [label="TotalCharges <= 3246.375\nentropy = 0.029\nsamples =
335\nvalue = [334, 1]\nclass = No"] ;
81 -> 82 ;
83 [label="entropy = 0.141\nsamples = 50\nvalue = [49, 1]\nclass = No"]
;
82 -> 83 ;
84 [label="entropy = 0.0\nsamples = 285\nvalue = [285, 0]\nclass = No"]
;
82 -> 84 ;
85 [label="tenure <= 71.5\nentropy = 0.195\nsamples = 133\nvalue =
[129, 4]\nclass = No"] ;
81 -> 85 ;
86 [label="entropy = 0.349\nsamples = 61\nvalue = [57, 4]\nclass = No"]
;
85 -> 86 ;
87 [label="entropy = 0.0\nsamples = 72\nvalue = [72, 0]\nclass = No"] ;
85 -> 87 ;
88 [label="tenure <= 26.5\nentropy = 0.906\nsamples = 1507\nvalue =
[485, 1022]\nclass = Yes"] ;
0 -> 88 [labeldistance=2.5, labelangle=-45, headlabel="False"] ;
89 [label="tenure <= 3.5\nentropy = 0.653\nsamples = 898\nvalue = [151,
747]\nclass = Yes"] ;
88 -> 89 ;
90 [label="TotalCharges <= 70.025\nentropy = 0.374\nsamples =
347\nvalue = [25, 322]\nclass = Yes"] ;
89 -> 90 ;
91 [label="PaperlessBilling_code <= 0.5\nentropy = 0.203\nsamples =
126\nvalue = [4, 122]\nclass = Yes"] ;
90 -> 91 ;
92 [label="entropy = 0.318\nsamples = 52\nvalue = [3, 49]\nclass =
Yes"] ;
91 -> 92 ;
93 [label="entropy = 0.103\nsamples = 74\nvalue = [1, 73]\nclass =
Yes"] ;
91 -> 93 ;
94 [label="MonthlyCharges <= 75.075\nentropy = 0.453\nsamples =
221\nvalue = [21, 200]\nclass = Yes"] ;
90 -> 94 ;
```

```
95 [label="entropy = 0.65\nsamples = 96\nvalue = [16, 80]\nclass =
Yes"] ;
94 -> 95 ;
96 [label="TotalCharges <= 162.025\nentropy = 0.242\nsamples =
125\nvalue = [5, 120]\nclass = Yes"] ;
94 -> 96 ;
97 [label="entropy = 0.0\nsamples = 69\nvalue = [0, 69]\nclass = Yes"]
;
96 -> 97 ;
98 [label="entropy = 0.434\nsamples = 56\nvalue = [5, 51]\nclass =
Yes"] ;
96 -> 98 ;
99 [label="MonthlyCharges <= 94.675\nentropy = 0.776\nsamples =
551\nvalue = [126, 425]\nclass = Yes"] ;
89 -> 99 ;
100 [label="OnlineSecurity_code <= 0.5\nentropy = 0.831\nsamples =
441\nvalue = [116, 325]\nclass = Yes"] ;
99 -> 100 ;
101 [label="MonthlyCharges <= 73.6\nentropy = 0.762\nsamples =
353\nvalue = [78, 275]\nclass = Yes"] ;
100 -> 101 ;
102 [label="MonthlyCharges <= 59.475\nentropy = 0.865\nsamples =
122\nvalue = [35, 87]\nclass = Yes"] ;
101 -> 102 ;
103 [label="entropy = 0.789\nsamples = 72\nvalue = [17, 55]\nclass =
Yes"] ;
102 -> 103 ;
104 [label="entropy = 0.943\nsamples = 50\nvalue = [18, 32]\nclass =
Yes"] ;
102 -> 104 ;
105 [label="TotalCharges <= 1277.3\nentropy = 0.693\nsamples =
231\nvalue = [43, 188]\nclass = Yes"] ;
101 -> 105 ;
106 [label="tenure <= 6.5\nentropy = 0.594\nsamples = 153\nvalue = [22,
131]\nclass = Yes"] ;
105 -> 106 ;
107 [label="entropy = 0.439\nsamples = 55\nvalue = [5, 50]\nclass =
Yes"] ;
106 -> 107 ;
108 [label="entropy = 0.666\nsamples = 98\nvalue = [17, 81]\nclass =
Yes"] ;
106 -> 108 ;
109 [label="entropy = 0.84\nsamples = 78\nvalue = [21, 57]\nclass =
Yes"] ;
105 -> 109 ;
110 [label="entropy = 0.987\nsamples = 88\nvalue = [38, 50]\nclass =
Yes"] ;
100 -> 110 ;
111 [label="TotalCharges <= 1173.675\nentropy = 0.439\nsamples =
110\nvalue = [10, 100]\nclass = Yes"] ;
99 -> 111 ;
112 [label="entropy = 0.0\nsamples = 50\nvalue = [0, 50]\nclass = Yes"]
;
111 -> 112 ;
113 [label="entropy = 0.65\nsamples = 60\nvalue = [10, 50]\nclass =
Yes"] ;
111 -> 113 ;
```

```
114 [label="Contract_code <= 1.5\nentropy = 0.993\nsamples = 609\nvalue
= [334, 275]\nclass = No"] ;
88 -> 114 ;
115 [label="tenure <= 59.5\nentropy = 0.998\nsamples = 483\nvalue =
[228, 255]\nclass = Yes"] ;
114 -> 115 ;
116 [label="MonthlyCharges <= 74.925\nentropy = 0.982\nsamples =
373\nvalue = [157, 216]\nclass = Yes"] ;
115 -> 116 ;
117 [label="MonthlyCharges <= 55.35\nentropy = 0.981\nsamples =
100\nvalue = [58, 42]\nclass = No"] ;
116 -> 117 ;
118 [label="entropy = 0.943\nsamples = 50\nvalue = [32, 18]\nclass =
No"] ;
117 -> 118 ;
119 [label="entropy = 0.999\nsamples = 50\nvalue = [26, 24]\nclass =
No"] ;
117 -> 119 ;
120 [label="MonthlyCharges <= 88.575\nentropy = 0.945\nsamples =
273\nvalue = [99, 174]\nclass = Yes"] ;
116 -> 120 ;
121 [label="entropy = 0.789\nsamples = 72\nvalue = [17, 55]\nclass =
Yes"] ;
120 -> 121 ;
122 [label="MonthlyCharges <= 94.775\nentropy = 0.975\nsamples =
201\nvalue = [82, 119]\nclass = Yes"] ;
120 -> 122 ;
123 [label="entropy = 0.99\nsamples = 52\nvalue = [29, 23]\nclass =
No"] ;
122 -> 123 ;
124 [label="TotalCharges <= 4199.975\nentropy = 0.939\nsamples =
149\nvalue = [53, 96]\nclass = Yes"] ;
122 -> 124 ;
125 [label="entropy = 0.989\nsamples = 64\nvalue = [28, 36]\nclass =
Yes"] ;
124 -> 125 ;
126 [label="entropy = 0.874\nsamples = 85\nvalue = [25, 60]\nclass =
Yes"] ;
124 -> 126 ;
127 [label="MonthlyCharges <= 99.125\nentropy = 0.938\nsamples =
110\nvalue = [71, 39]\nclass = No"] ;
115 -> 127 ;
128 [label="entropy = 0.804\nsamples = 57\nvalue = [43, 14]\nclass =
No"] ;
127 -> 128 ;
129 [label="entropy = 0.998\nsamples = 53\nvalue = [28, 25]\nclass =
No"] ;
127 -> 129 ;
130 [label="MonthlyCharges <= 102.775\nentropy = 0.631\nsamples =
126\nvalue = [106, 20]\nclass = No"] ;
114 -> 130 ;
131 [label="entropy = 0.398\nsamples = 76\nvalue = [70, 6]\nclass =
No"] ;
130 -> 131 ;
132 [label="entropy = 0.855\nsamples = 50\nvalue = [36, 14]\nclass =
No"] ;
130 -> 132 ;
}
```

Rahul Kumar