

Telecom Churn Case Study

Business Problem

- In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, **customer retention** has now become even more important than customer acquisition.
- For many incumbent operators, *retaining high profitable customers is the number one business goal.*
- To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn.**

PROBLEM DEFINITION

“To predict the customers who are going to churn”

Basically, it is a binary classification problem where the output will be 1 if it is going to be churned otherwise 0 if not going to be churned

Filtering high value customers

After Filtering high value customers, we got around (30011, 179)

Here we can see that we got approx 30K rows

Identifying Churn

0 91.359835

1 8.640165

Training and testing the data

DECISION TREES

Decision Tree Report with PCA:

	precision	recall	f1-score	support	Training Data Metrics:
0	0.94	0.93	0.94	5476	Accuracy: 1.00
1	0.35	0.37	0.36	527	Sensitivity (Recall): 1.00
					Specificity: 1.00
accuracy			0.88	6003	Test Data Metrics:
macro avg	0.64	0.65	0.65	6003	Accuracy: 0.88
weighted avg	0.89	0.88	0.89	6003	Sensitivity (Recall): 0.37
					Specificity: 0.93

Random Forest

Random Forest Report with PCA:					Random Forest Training Data Metrics:	
	precision	recall	f1-score	support	Accuracy:	1.00
0	0.93	0.98	0.96	5476	Sensitivity (Recall):	1.00
1	0.63	0.29	0.39	527	Specificity:	1.00
					Random Forest Test Data Metrics:	
accuracy			0.92	6003	Accuracy:	0.92
macro avg	0.78	0.64	0.68	6003	Sensitivity (Recall):	0.29
weighted avg	0.91	0.92	0.91	6003	Specificity:	0.98

Insights

- The Random Forest model has a high accuracy and specificity on the training data, suggesting that it's doing an excellent job of identifying non-churn cases.
- On the test data, the Random Forest model maintains a reasonably high accuracy and specificity, which indicates its ability to generalize well to unseen data.
- The sensitivity (recall) for both models is low, especially on the test data. This implies that both models struggle to correctly identify customers who are likely to churn (churn=1). It's essential to improve this aspect of the models to reduce false negatives.
- The specificity is relatively high, indicating that the models are good at identifying non-churn cases. This is important for not mistakenly categorizing loyal customers as churners.

Insights

- The Random Forest model performs significantly better than the Decision Tree model on both training and test data.
- It has a higher accuracy and specificity, which indicates that the Random Forest model is better at correctly classifying non-churn (0) cases. This is important for avoiding false alarms and not misclassifying loyal customers.
- The Decision Tree model appears to be overfitting the training data, as evidenced by the perfect accuracy, sensitivity, and specificity on the training data but lower performance on the test data. Overfitting means the model has learned the training data too well and struggles to generalize to unseen data.

Columns in the dataframe

```
Index(['mobile_number', 'arpu_6', 'arpu_7', 'arpu_8', 'total_og_mou_6',  
      'total_og_mou_7', 'total_og_mou_8', 'total_ic_mou_6', 'total_ic_mou_7',  
      'total_ic_mou_8', 'total_rech_num_6', 'total_rech_num_7',  
      'total_rech_num_8', 'total_rech_amt_6', 'total_rech_amt_7',  
      'total_rech_amt_8', 'max_rech_amt_6', 'max_rech_amt_7',  
      'max_rech_amt_8', 'last_day_rch_amt_6', 'last_day_rch_amt_7',  
      'last_day_rch_amt_8', 'vol_2g_mb_6', 'vol_2g_mb_7', 'vol_2g_mb_8',  
      'vol_3g_mb_6', 'vol_3g_mb_7', 'vol_3g_mb_8', 'monthly_2g_6',  
      'monthly_2g_7', 'monthly_2g_8', 'sachet_2g_6', 'sachet_2g_7',  
      'sachet_2g_8', 'monthly_3g_6', 'monthly_3g_7', 'monthly_3g_8',  
      'sachet_3g_6', 'sachet_3g_7', 'sachet_3g_8', 'aon', 'aug_vbc_3g',  
      'jul_vbc_3g', 'jun_vbc_3g', 'sep_vbc_3g', 'avg_rech_amt_6_7', 'churn'],  
      dtype='object')
```


Top 10 Features

	Feature	Importance
5	total_og_mou_7	0.158805
1	arpu_6	0.143309
7	total_ic_mou_6	0.105981
8	total_ic_mou_7	0.095118
4	total_og_mou_6	0.092979
6	total_og_mou_8	0.084929
3	arpu_8	0.084453
2	arpu_7	0.082973
9	total_ic_mou_8	0.079465
0	mobile_number	0.071988

