Telecom Churn

Business Analysis Case Study -1 Assignment

By Monica Kamalakar

Problem statement:-

To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.

In this project, we will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

Retaining high profitable customers is the main business goal here.

Solution Plan of Action:

1. Reading, understanding and visualising the data
2. Preparing the data for modelling
3. Building the model
4. Evaluate the model

Steps:

Reading and understanding the data

Handling missing values

Deleting the date columns as the date columns are not required in our analysis

Filter high-value customers

Handling missing values in rows¶

Outliers treatment

Derive new features

EDA

Univariate analysis

Model with PCA

Logistic regression with PCA

Build the model with optimal hyperparameters

Prediction on the train set

Decision tree with PCA

Prediction on the test set

Without PCA

Logistic regression with No PCA¶

Feature Selection Using RFE -RFE with 15 columns

Checking VIFs

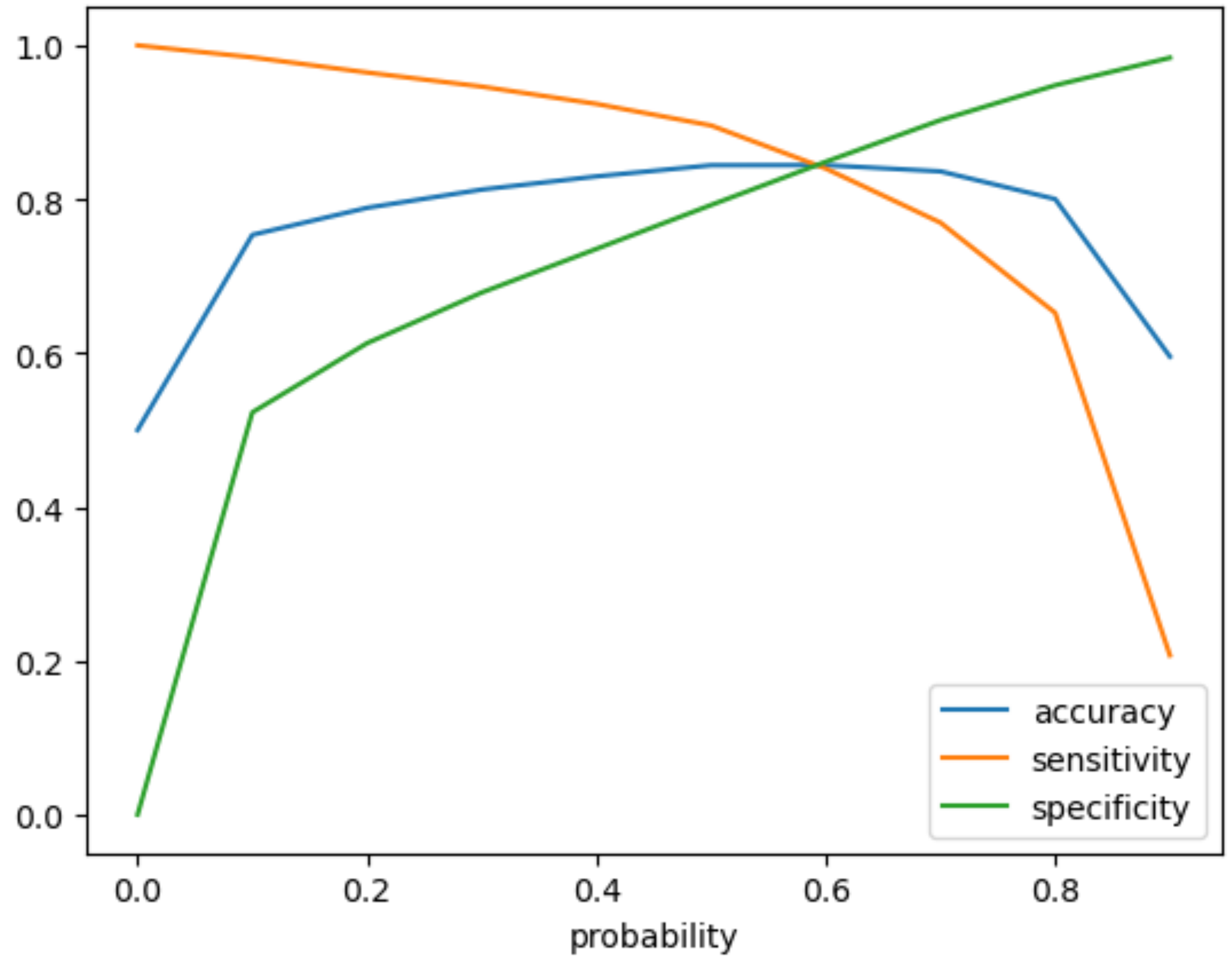
Business recomendation

Analysis

- From the model summary and the VIF list we can see that all the variables are significant and there is no multicollinearity among the variables.
- Hence, we can conclude that ***Model-3 lg_3*** *will be the final model*.

Accuracy, Sensitivity and Specificity curve

- Accuracy - Becomes stable around 0.6
- Sensitivity - Decreases with the increased probability.
- Specificity - Increases with the increasing probability.
- At point 0.6 where the three parameters cut each other, we can see that there is a balance between sensitivity and specificity with a good accuracy.
- Here we are intended to achieve better sensitivity than accuracy and specificity. Though as per the above curve, we should take 0.6 as the optimum probability cutoff, we are taking *0.5* for achieving higher sensitivity, which is our main goal.



Model summary

- Train set
  - Accuracy = 0.84
  - Sensitivity = 0.81
  - Specificity = 0.83
- Test set
  - Accuracy = 0.78
  - Sensitivity = 0.82
  - Specificity = 0.78

Overall, the model is performing well in the test set, what it had learnt from the train set.

Business recommendation

Top predictorschurn

Below are few top variables selected in the logistic regression model.

We can see most of the top variables have negative coefficients. That means, the variables are inversely correl
the churn probablity.

E.g.:-

If the local incoming minutes of usage (loc_ic_mou_8) is lesser in the month of August than any other month
there is a higher chance that the customer is likely to.

Variables	Coefficients
loc_ic_mou_8	-3.3287
og_others_7	-2.4711
ic_others_8	-1.5131
isd_og_mou_8	-1.3811
decrease_vbc_action	-1.3293
monthly_3g_8	-1.0943
std_ic_t2f_mou_8	-0.9503
monthly_2g_8	-0.9279
loc_ic_t2f_mou_8	-0.7102
roam_og_mou_8	0.7135

Recomendations

1. Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
2. Target the customers, whose outgoing others charge in July and incoming others on August are less.
3. Also, the customers having value based cost in the action phase increased are more likely to churn than the other customers. Hence, these customers may be a good target to provide offer.
4. Cutomers, whose monthly 3G recharge in August is more, are likely to be churned.
5. Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
6. Cutomers decreasing monthly 2g usage for August are most probable to churn.
7. Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.
8. roam_og_mou_8 variables have positive coefficients (0.7135). That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.

THANK YOU