Telecom-Churn-Case-study---IIITB-Assignment

MHONISHA N

Git Hub Links

- https://github.com/MhonishaN/Telecom-Churn
- telecom-churn-case-study (Mhonisha).ipynb

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.
Steps:-
Reading, understanding and visualising the data
Preparing the data for modelling
Building the model
Evaluate the model

Final conclusion with PCA
After trying several models we can see that for achieving the best sensitivity, which was our ultimate goal, the classic Logistic regression or the SVM models preforms well. For both the models the sensitivity was approx. 81%. Also we have good accuracy of approx. 85%.

Final conclusion with no PCA
We can see that the logistic model with no PCA has good sensitivity and accuracy, which are comparable to the models with PCA. So, we can go for the more simplistic model such as logistic regression with PCA as it explains the important predictor variables as well as the significance of each variable. The model also helps us to identify the variables which should be act upon for making the decision of the to be churned customers. Hence, the model is more relevant in terms of explaining to the business.

Business recommendation: Top predictors

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

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

We can see most of the top variables have negative coefficients. That means, the variables are inversely correlated with the churn probability.
E.g.:-
If the local incoming minutes of usage (loc_ic_mou_8) is lesser in the month of August than any other month, then there is a higher chance that the customer is likely to churn.

Recommendations

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.Customers, whose monthly 3G recharge in August is more, are likely to be churned.

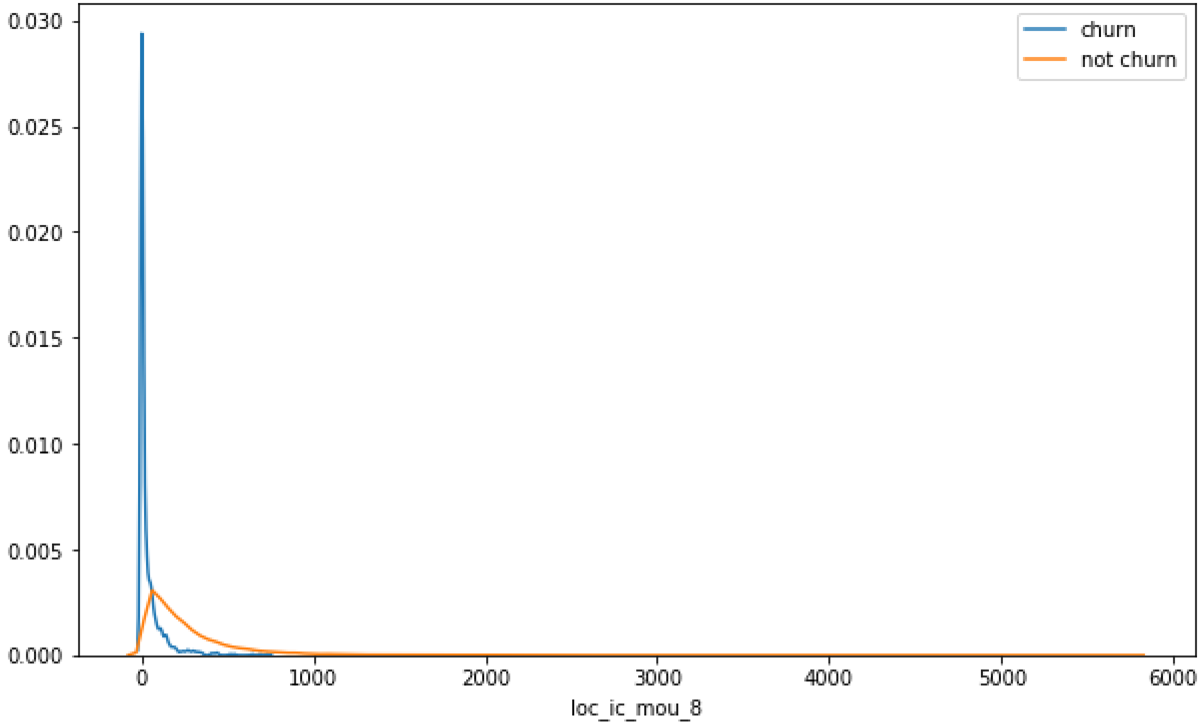
5.Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.

6.Customers 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.

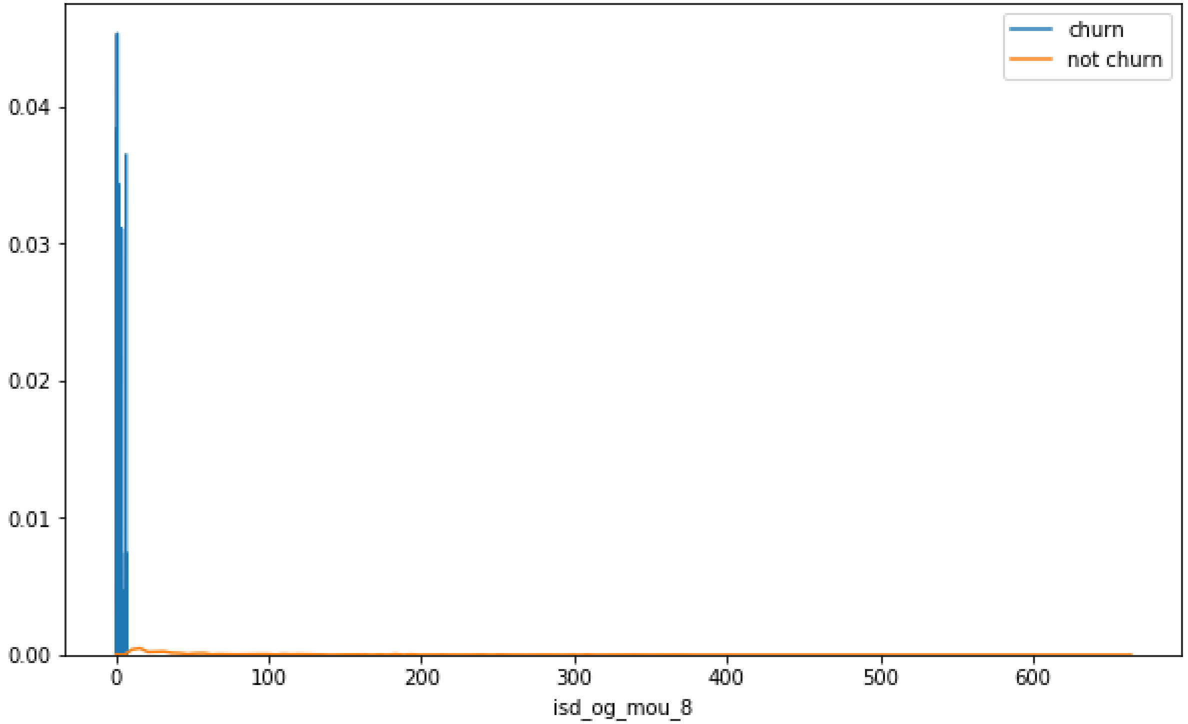
8.roam_og_mou_8 variables have positive coefficients (0.7135). That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.

Plots of important predictors for churn and non churn customers



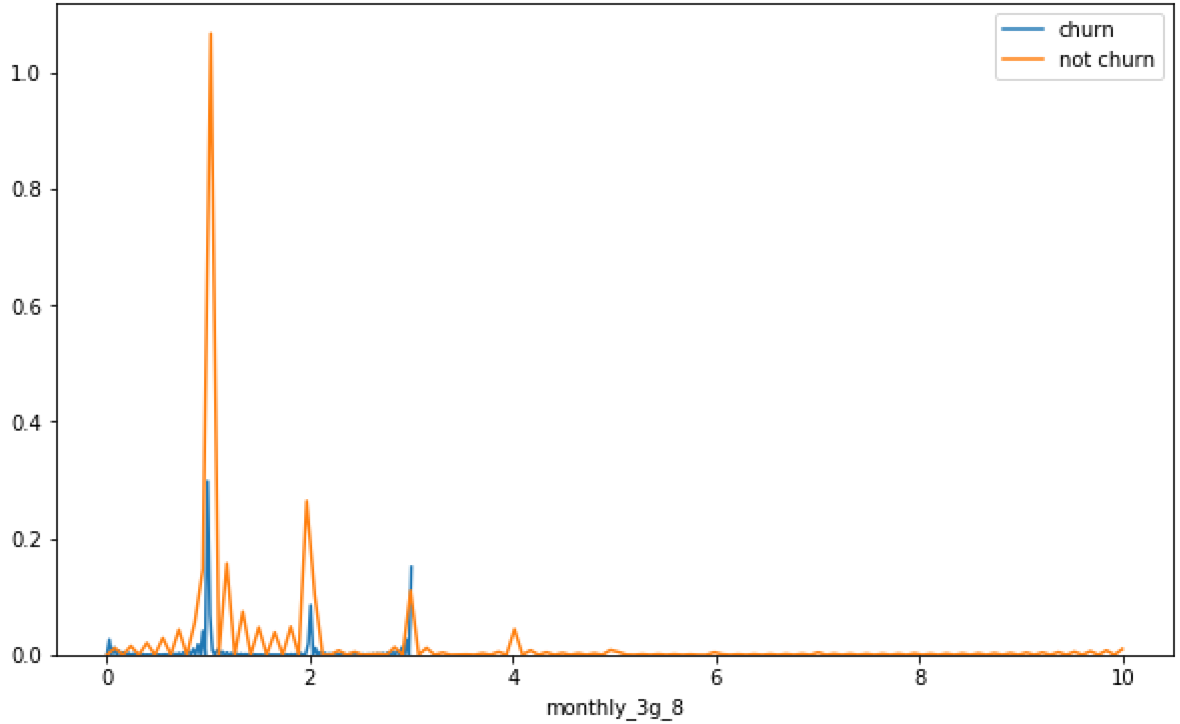
We can see that for the churn customers the minutes of usage for the month of August is mostly populated on the lower side than the non churn customers.

Plots of important predictors for churn and non churn customers



We can see that the ISD outgoing minutes of usage for the month of August for churn customers is densed approximately to zero. On the onther hand for the non churn customers it is little more than the churn customers.

Plots of important predictors for churn and non churn customers



The number of mothly 3g data for August for the churn customers are very much populated aroud 1, whereas of non churn customers it spreaded accross various numbers.

Similarly we can plot each variables, which have higher coefficients, churn distribution.

Overall Conclusions

- Std Outgoing Calls and Revenue Per Customer are strong indicators of Churn.
- Local Incoming and Outgoing Calls for 8th Month and avg revenue in 8th Month are the most important columns to predict churn.
- Customers with tenure less than 4 yrs are more likely to churn.
- Max Recharge Amount is a strong feature to predict churn. Random Forest produced the best prediction results followed by SVM.

Business Insights

Telecom company needs to pay attention to the roaming rates.
They need to provide good offers to the customers who are using services from a roaming zone.
The company needs to focus on the STD and ISD rates. Perhaps, the rates are too high.
Provide them with some kind of STD and ISD packages.
To look into both of the issues stated above, it is desired that the telecom company collects customer query and complaint data and work on their services according to the needs of customers.