TELECOM CHURN CASE STUDY

Introduction to Telecom Churn

1. Definition of Telecom Churn

- Telecom Churn, also known as customer churn or attrition, refers to the phenomenon where customers terminate their subscription or services with a telecom service provider. It signifies the loss of customers to competing telecom companies or other factors.

Significance of Studying Churn in the Telecom Industry

- Churn is a critical metric for telecom companies as it directly impacts their revenue and profitability.
- Understanding churn patterns allows telecom providers to identify the reasons why customers leave, enabling them to take proactive measures to retain customers.
- Reducing churn not only preserves the customer base but also reduces acquisition costs associated with acquiring new customers.
- Churn analysis provides valuable insights into customer preferences, behavior, and market dynamics, which can guide product development and marketing strategies.

Analysis

As expected, the churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase.

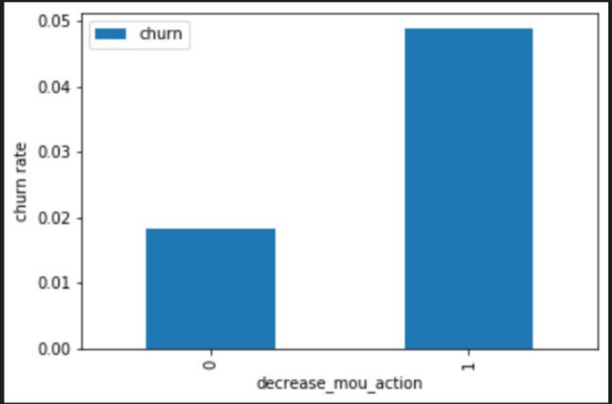
Churn rate on the basis whether the customer decreased her/his MOU in action month

```
# Converting churn column to int in order to do aggfunc in the pivot table
data['churn'] = data['churn'].astype('int64')
```

[75]

```
data.pivot_table(values='churn', index='decrease_mou_action', aggfunc='mean').plot.bar()
plt.ylabel('churn rate')
plt.show()
```

[76]



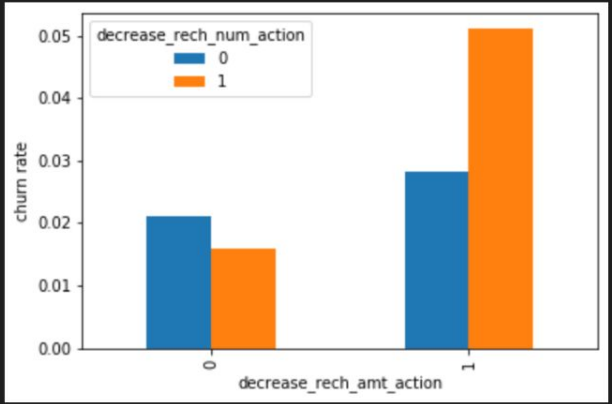
Bivariate analysis

Analysis of churn rate by the decreasing recharge amount and number of recharge in the action phase

```
data.pivot_table(values='churn', index='decrease_rech_amt_action', columns='decrease_rech_num_action', aggfunc='mean').plot.bar()
plt.ylabel('churn rate')
plt.show()
```

[83]

...

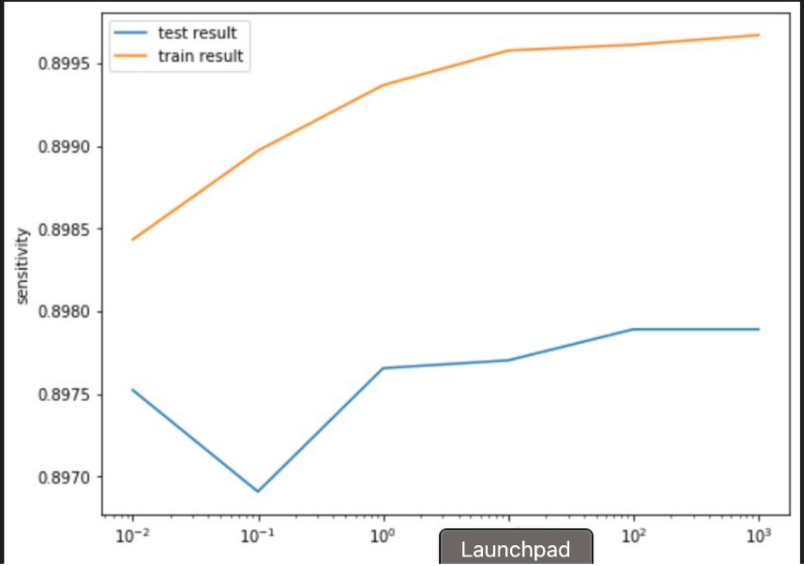


Analysis

```python
# plot of C versus train and validation scores

plt.figure(figsize=(8, 6))
plt.plot(cv_results['param_C'], cv_results['mean_test_score'])
plt.plot(cv_results['param_C'], cv_results['mean_train_score'])
plt.xlabel('C')
plt.ylabel('sensitivity')
plt.legend(['test result', 'train result'], loc='upper left')
plt.xscale('log')
```

[117]



Model-1 with optimal hyperparameters

```
Accuracy:- 0.8012994044396319
Sensitivity:- 0.7564766839378239
Specificity:- 0.8029169783096485
```

Model summary

- Train set
  - Accuracy = 0.84
  - Sensitivity = 0.88
  - Specificity = 0.80
- Test set
  - Accuracy = 0.80
  - Sensitivity = 0.75
  - Specificity = 0.80

We can see from the model performance that the Sesitivity has been decreased while evaluating the model on the test set. However, the accuracy and specificity is quite good in the test set.

Final conclusion with PCA

After trying several models we can see that for acheiving 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 apporx 85%.

MODEL- 2 Checking VIF for Model-2

	Features	VIF
1	offnet_mou_8	7.45
3	std_og_t2m_mou_8	6.27
0	offnet_mou_7	1.92
7	loc_ic_mou_8	1.68
6	loc_ic_t2f_mou_8	1.21
10	total_rech_num_8	1.19
2	roam_og_mou_8	1.16
13	decrease_vbc_action	1.08
12	monthly_3g_8	1.06
11	monthly_2g_8	1.05
8	std_ic_t2f_mou_8	1.02
4	isd_og_mou_8	1.01
9	ic_others_8	1.01
5	og_others_7	1.00

As we can see from the model summary that all the variables p-values are significant and offnet_mou_8 column has the highest VIF 7.45. Hence, deleting offnet_mou_8 column.

VIF Model-3

```
Accuracy:- 0.7848763761053962
Sensitivity:- 0.8238341968911918
Specificity:- 0.7834704562453254
```

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.

Final Conclusion

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.

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 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, 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. 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.

THANKYOU