

Telecom Churn Casestudy

By:

Surya Prabha

Sharath Nair

Abhishek Kaushik

Contents

- Problem Statement
- High Value Customer
- Tagging Churn
- Data Cleaning & EDA
- Data Preparation
- Model Building
- Conclusions

Problem Statement

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.

In this project, you 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.

High Value Customer

The screenshot displays a web browser with multiple tabs open, including 'upGrad | Learning Platform', 'GitHub', and several instances of 'Telecom Churn Case Study'. The active page is a Kaggle notebook titled 'Telecom Churn Case Study - UpGrad'. The notebook interface includes a search bar, 'Sign In', and 'Register' buttons. A left sidebar shows navigation options: 'Create', 'Home', 'Competitions', 'Datasets', 'Models', 'Code' (selected), 'Discussions', 'Learn', and 'More'. The main area shows a code cell with the following Python code:

```
In [7]:  
# Though we have total amount of recharge for talktime we dont have that number for data, instead  
# we have average and number of recharges  
# so let's calculate that amount  
churn_data['total_data_rech_amt_6'] = churn_data['av_rech_amt_data_6'] * churn_data['total_rech_data_6']  
churn_data['total_data_rech_amt_7'] = churn_data['av_rech_amt_data_7'] * churn_data['total_rech_data_7']  
churn_data['total_data_rech_amt_8'] = churn_data['av_rech_amt_data_8'] * churn_data['total_rech_data_8']  
churn_data['total_data_rech_amt_9'] = churn_data['av_rech_amt_data_9'] * churn_data['total_rech_data_9']  
  
# Drop the columns already accounted for  
churn_data.drop(['total_rech_data_6', 'total_rech_data_7', 'total_rech_data_8', 'total_rech_data_9',  
                'av_rech_amt_data_6',  
                'av_rech_amt_data_7', 'av_rech_amt_data_8', 'av_rech_amt_data_9'], axis=1, inplace=True)  
  
# Find Average recharge amount for good phase i.e 6th and 7th month
```

On the right side, a 'Table of Contents' panel lists the notebook's sections: 'High Value Customer', 'Tagging Churn', 'Data Cleaning & EDA', 'Data Preparation', 'Model Building', and 'Conclusions'. At the bottom of the browser window, a Windows taskbar is visible with the date '7/11/2023' and time '11:48 PM'. A file explorer bar at the bottom of the notebook shows open files: 'Telecom_Churn_...ipynb', 'telecom_churn_data.csv', 'Data+Dictionary-...xlsx', and 'README.md'.

Tagging Churn

The screenshot displays a Kaggle notebook interface for a project titled "Telecom Churn Case Study - UpGrad". The notebook is in "Code" view, showing two code cells. The first cell calculates the churn rate for each customer, and the second cell drops specific columns from the dataset. The output of the first cell shows churn rates for churn status 0 and 1. The output of the second cell shows the dimensions of the dataset after dropping columns.

Code Cell 12:

```
In [12]: churn_data.churn.value_counts(normalize=True)
```

Output 12:

```
0    0.942369
1    0.057631
Name: churn, dtype: float64
```

Code Cell 13:

```
In [13]: # Dropping Month 9 Columns

churn_data.drop( [ col for col in mnth9_columns + churn_tag_columns if col not in ['total_rech_d
ata_9', 'av_rech_amt_data_9'] ],
                axis=1, inplace=True)
churn_data.shape
```

Output 13:

```
(30001, 171)
```

Table of Contents:

- High Value Customer
- Tagging Churn
- Data Cleaning & EDA
- Data Preparation
- Model Building
- Conclusions

File Explorer:

- Telecom_Churn_....ipynb
- telecom_churn_data.csv
- Data+Dictionary-....xlsx
- README.md

System Tray: 84°F, Mostly cloudy, 11:49 PM, 7/11/2023

EDA

The screenshot shows a web browser with multiple tabs open, including 'upGrad | Learning Platform', 'GitHub', and several instances of 'Telecom Churn Case Study'. The active page is a Kaggle notebook titled 'Telecom Churn Case Study - UpGrad' by user 'spramanik'. The notebook is in 'Code' view and shows a Jupyter cell with the following code and output:

```
In [14]: # Checking for missing values
mmsg_values = round(churn_data.isna().sum() / len(churn_data) * 100, 2).sort_values(ascending=False)
mmsg_values
```

The output of the cell is a list of features and their corresponding percentage of missing values, sorted in descending order:

Feature	Percentage of Missing Values
total_data_rech_amt_9	49.27
arpu_3g_8	46.83
total_data_rech_amt_8	46.83
fb_user_8	46.83
max_rech_data_8	46.83
count_rech_2g_8	46.83
count_rech_3g_8	46.83
night_pck_user_8	46.83
date_of_last_rech_data_8	46.83
arpu_2g_8	46.83
arpu_3g_6	44.15

The right sidebar of the notebook shows a 'Table of Contents' with links to various sections: 'High Value Customer', 'Tagging Churn', 'Data Cleaning & EDA' (which is the current section), 'Data Preparation', 'Model Building', and 'Conclusions'. The bottom of the screen shows a Windows taskbar with the date '7/11/2023' and time '11:49 PM'.

Data Prep

The screenshot shows a web browser with multiple tabs open, including 'upGrad | Learning Platform', 'GitHub', and several 'Telecom Churn Case Study' tabs. The active page is a Kaggle notebook titled 'Telecom Churn Case Study - UpGrad'. The notebook interface includes a search bar, 'Sign In', and 'Register' buttons. The left sidebar shows navigation options: Home, Competitions, Datasets, Models, Code (selected), Discussions, Learn, and More. The main content area displays the notebook's title, tabs for Notebook, Input, Output, Logs, and Comments (0), and a 'Copy & Edit' button. The notebook content shows a code cell with the following text: 'Now the class is balanced and the target variable is not skewed', followed by a section header 'PCA'. Below this, the code cell shows the execution of 'X.shape' resulting in '(28163, 55)', and then the execution of PCA code resulting in '(54590, 25)'. A 'Table of Contents' on the right lists: High Value Customer, Tagging Churn, Data Cleaning & EDA, Data Preparation (highlighted), Model Building, and Conclusions. At the bottom, a cookie notice states: 'We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies.' Below the notice are tabs for 'Telecom_Churn_...ipynb', 'telecom_churn_data.csv', 'Data+Dictionary-...xlsx', and 'README.md'. The Windows taskbar at the bottom shows the date and time as '11:50 PM 7/11/2023'.

upGrad | Learning Platform x GitHub x Fwd: Churn Case Study - su x Telecom-Churn-Case-Study x Telecom Churn Case Study x Telecom Churn Case Study x +

kaggle.com/code/spramanik/telecom-churn-case-study-upgrad#Data-Preparation

Search

Sign In Register

Telecom Churn Case Study - UpGrad

Notebook Input Output Logs Comments (0)

5 Copy & Edit 77

Now the class is balanced and the target variable is not skewed

PCA

In [48]:

```
X.shape
```

Out[48]:

```
(28163, 55)
```

In [49]:

```
from sklearn.decomposition import PCA

pca = PCA(n_components=25)
X_pca = pca.fit_transform(X_res)
X_pca.shape
```

Out[49]:

```
(54590, 25)
```

Table of Contents

- High Value Customer
- Tagging Churn
- Data Cleaning & EDA
- Data Preparation**
- Model Building
- Conclusions

We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies. Got it Learn more

Telecom_Churn_...ipynb telecom_churn_data.csv Data+Dictionary-...xlsx README.md Show all x

84°F Mostly cloudy Search 11:50 PM 7/11/2023

Model Building

The screenshot displays a Kaggle notebook interface for a 'Telecom Churn Case Study - UpGrad'. The browser tabs at the top include 'upGrad | Learning Platform', 'GitHub', 'Fwd: Churn Case Study - su...', 'Telecom-Churn-Case-Study', and several instances of 'Telecom Churn Case Study'. The URL in the address bar is 'kaggle.com/code/spramanik/telecom-churn-case-study-upgrad#Model-Building'. The notebook's left sidebar shows navigation options: 'Create', 'Home', 'Competitions', 'Datasets', 'Models', 'Code' (selected), 'Discussions', 'Learn', and 'More'. The main content area is titled 'Telecom Churn Case Study - UpGrad' and includes tabs for 'Notebook', 'Input', 'Output', 'Logs', and 'Comments (0)'. It shows '5' versions and a 'Copy & Edit' button with a count of '77'. The current section is '1. Logistic Regression', which includes a text block stating: 'For logistic regression we will be using the unaltered X and y so that we can use RFE for feature selection instead of PCA, to find out the strong predictor of churn'. Below this are three code cells:
Cell 50:

```
# Split the data into train and test
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_res, y_res, train_size=0.7, test_size=0.3, random_state=25)
```


Cell 51:

```
import statsmodels.api as sm
```


Cell 52:

```
# Initial logistic regression model
logm1 = sm.GLM(y_train, (sm.add_constant(X_train)), family = sm.families.Binomial())
```


A 'Table of Contents' on the right lists: 'High Value Customer', 'Tagging Churn', 'Data Cleaning & EDA', 'Data Preparation', 'Model Building' (highlighted), and 'Conclusions'. A cookie notice at the bottom states: 'We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies.' The taskbar at the very bottom shows the system clock as 11:50 PM on 7/11/2023.

upGrad | Learning Platform

GitHub

Fwd: Churn Case Study - su

Telecom-Churn-Case-Study

Telecom Churn Case Study

Telecom Churn Case Study

+

←

→

↻

🏠

kaggle.com/code/spramanik/telecom-churn-case-study-upgrad#Conclusions

📄

🔖

☆

⚙️

🖥️

👤

⋮

≡

kaggle

+

 Create

🏠 Home

🏆 Competitions

📁 Datasets

👤 Models

<>

 Code

💬 Discussions

🎓 Learn

⌵ More

🔍 Search

Sign In

Register

Telecom Churn Case Study - UpGrad

Notebook

Input

Output

Logs

Comments (0)

⬆

5

Copy & Edit

77

⋮

In [108]:

```
# Plot the scores corresponding to number of estimators to find the best possible number of estimator
plt.plot(estimators, adaboost_scores)
plt.xlabel('n_estimators')
plt.ylabel('accuracy')
plt.show()
```



Table of Contents

>

High Value Customer

Tagging Churn

Data Cleaning & EDA

Data Preperation

Model Building

Conclusions

We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies.

Got it

Learn more

Telecom_Churn_...ipynb

telecom_churn_data.csv

Data+Dictionary-...xlsx

README.md

Show all

×

84°F Mostly cloudy

🏠 🔍 Search 🌧️ 🖨️ 📁 🗂️ 🌐 🔄 📶 🔊 🔋

11:51 PM 7/11/2023 🔔

Conclusion

The screenshot shows a Kaggle notebook interface. The browser tabs at the top include 'upGrad | Learning Platform', 'GitHub', 'Fwd: Churn Case Study - stu', 'Telecom-Churn-Case-Study', and several instances of 'Telecom Churn Case Study'. The address bar shows the URL: `kaggle.com/code/spramanik/telecom-churn-case-study-upgrad#Conclusions`.

The notebook title is 'Telecom Churn Case Study - UpGrad'. The left sidebar shows navigation options: Home, Competitions, Datasets, Models, Code (selected), Discussions, Learn, and More.

The notebook content is titled 'Strategies to Manage Customer Churn'. It states: 'The top 10 predictors are :'

Features
loc_og_mou_8
total_rech_num_8
monthly_3g_8
monthly_2g_8
gd_ph_loc_og_mou
gd_ph_total_rech_num
last_day_rch_amt_8
std_ic_t2t_mou_8
sachet_2g_8
aon

Below the table, a bullet point states: 'We can see most of the top predictors are from the action phase, as the drop in engagement is prominent in that phase'.

Text: 'Some of the factors we noticed while performing EDA which can be clubbed with these insights are:'

1. Users whose maximum recharge amount is less than 200 even in the good phase, should have a tag and re-evaluated time to time as they are more likely to churn

On the right, a 'Table of Contents' sidebar lists: High Value Customer, Tagging Churn, Data Cleaning & EDA, Data Preparation, Model Building, and Conclusions (which is the active section).

At the bottom, a cookie notice states: 'We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies.' Buttons for 'Got it' and 'Learn more' are present.

The bottom of the screen shows a Windows taskbar with the date '7/11/2023' and time '11:52 PM'.