upGrad

#LifeKoKaroLift

Telecom Churn Case Study Doubts Session


upGrad
Course : Data Science
Lecture On : Assignment
Instructor : Sumit Shukla

upGrad

What we will cover in this session?

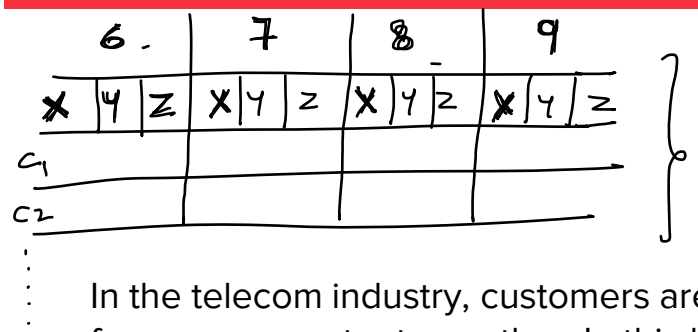
1. Case Study Walkthrough
2. QnA

Assignment

customer usage data

NO ready made target column provided.

upGrad



Problem Statement

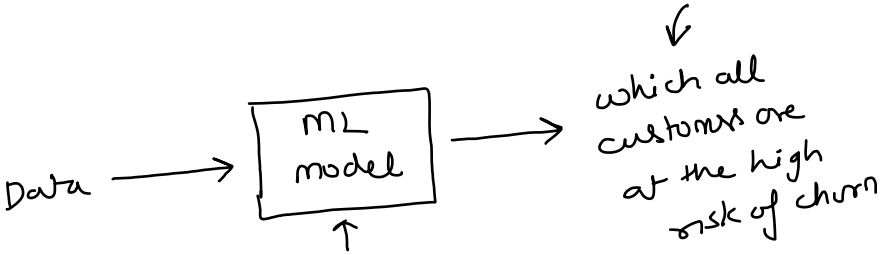
6, 7 → Good phase ← customers happy with services

8 → Action phase ← started facing some issues

9 → churn phase ←

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.

Predict which customers are at high risk of churn.



Assignment

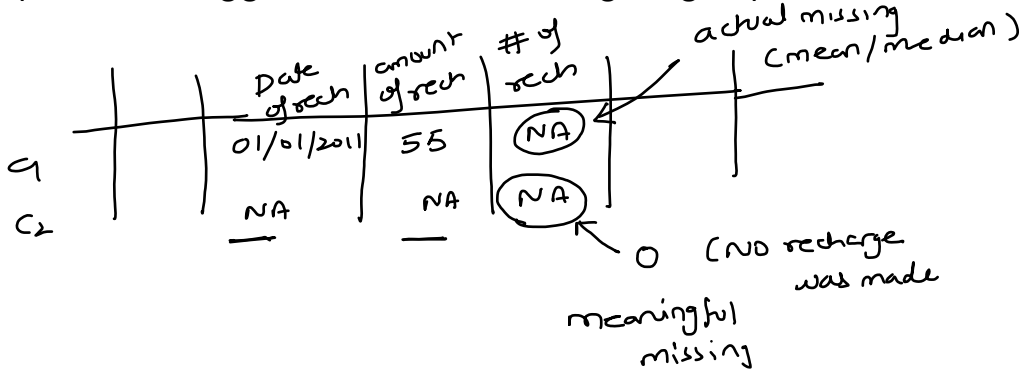
upGrad

Data cleaning

What you need to do?

1. Handling Missing data.
   a. Impute with zero when you are very sure that a missing is a zero.
   b. For categorical, what to do?
   c. Remove those with high missing percentage.

https://www.kaggle.com/athi94/investigating-imputation-methods



i) find and impute those rows and columns where a missing can be imputed with zero

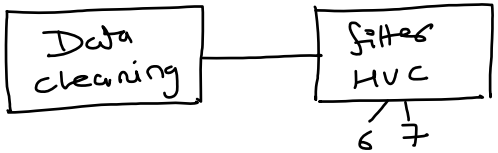
ii) find out percentage of missing data for various columns and drop those columns with high %age of missing

iii) Impute columns with less percentage of missing

mean / median / mode

Assignment

upGrad



Filter high-value customers(HVC)

80/20 Pareto rule : 80% Revenue → top 20% customers

↳ top 30%

Good Phase

1. Calculate average recharge done by customer in June and July(total_rech_amt)
2. Look at the 70th percentile recharge amount
3. Retain only those customers who have recharged their mobiles with more than or equal to 70th percentile amount

total rech-amt 6	total rech-amt 7	avg-rech amt-6-7
a	b	(a+b)/2

i) Create new column avg-rech-amt-6-7 which is the avg of total-rech-amt-6 & total-rech-amt-7

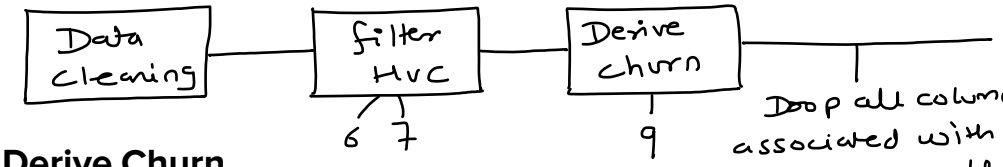
ii) find the 70th percentile value of for avg-rech-amt-6-7

iii) If avg-rech-amt-6-7 for a customer greater than or equal to 70th percentile value

fine { 30.1k, 29.9k, 29.8k } HVC ← { avg-rech amt-6-7 ≥ 70th percentile } ↙ HVC

Assignment

upGrad



Derive Churn

9th Month is our Churn Phase. Usage-based churn

1. Calculate total incoming and outgoing minutes of usage
2. Calculate 2g and 3g data consumption
3. Create churn variable: those who have not used either calls or internet in the month of September are customers who have churned
4. Check Churn percentage.
5. Delete columns that belong to the churn month

If for month 9

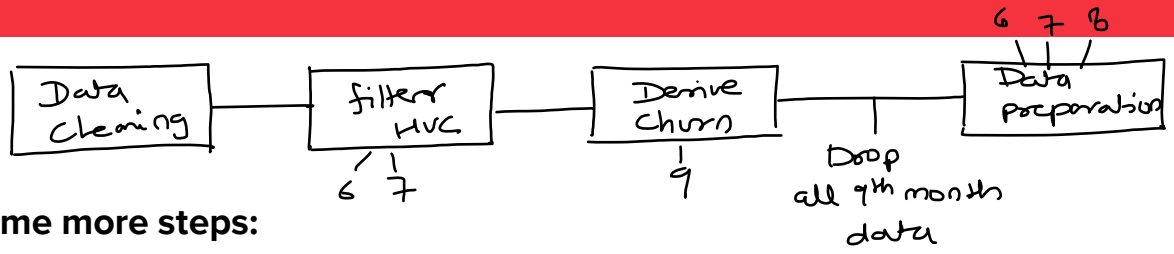
$y = 1$ else $y = 0$

total_incoming_mou == 0
&
total_outgoing_mou == 0
&
total_3g_data == 0
&
total_2g_data == 0

churn
- usage based churn → customer is not at all active for any service
- Revenue based churn → using incoming & SMS service but not making any recharge / generating any revenue for the company

Assignment

upGrad

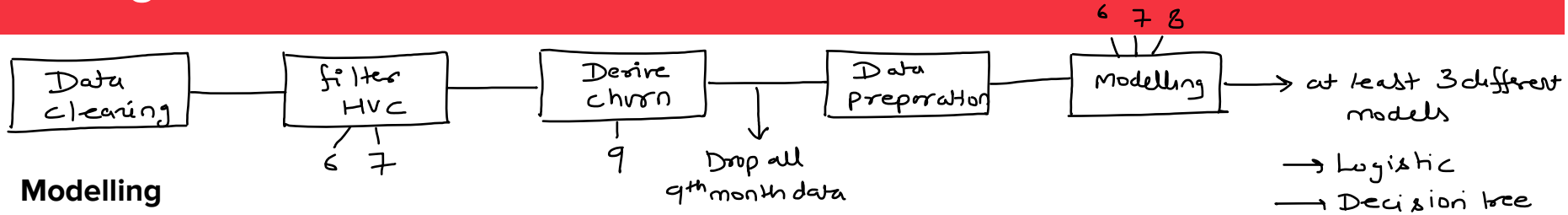


Some more steps:

1. Derived variable. at least 3 additional
2. EDA
3. Outlier treatment
4. split train-test
5. Scaling

Assignment

upGrad



Modelling

1. Try out various models and select the best among them.
2. You need to handle the imbalance class for all the models.

https://towardsdatascience.com/methods-for-dealing-with-imbalanced-data-5b761be45a18

within Jupyter notebook & markdown comment

using the logistic Regression model add recommendations for the company so that they can may avoid churn

upGrad

#LifeKoKaroLift