EXPLORE THE DATASET :-

The dataset contains data for June, July, August, September encoded as 6,7,8,9 respectively

The task is to predict the CHURN in the September based on Data From June, July, August

June, July - customer's happy phase
August - the action phase
September - customer's churn phase

PREPARE THE DATASET :-

Derive new FEATURES

Filter high values customers

Tag churners based on Sept. data

PREPARE THE MODELS :-

One model to predict whether high value customer will churn or not

Another to find strong predictor variables responsible for CHURN

create a feature called High_value_customer which takes 1 for true else 0, put values in this feature if rech_amount >= 70th percentile of (total_rech_amt_6+total_rech_amt_7)/2

circle_id can be discarded because it has only one value (109), number can be discarded, loc_og_t2o_mou, std_og_t2o_mou, can also be discarded because it has only 0s and NaN, last_date_of_month_6, last_date_of_month_7, last_date_of_month_8, last_date_of_month_9 can also be discarded

Those who have recharged with amount higher than or equal to 70th percentile of average recharge amount in customer's happy phase

CHURN=1, NON-CHURN=0 based on customer's CHURN phase by checking if they had made any calls or have they used Internet in the CHURN phase. After data labelling discard all features of September

customer will churn or not

Churn to be considered :- usage based churn

Predict Churn on :- High Value Customers

Handle Class Imbalance :- Since CHURN is typically low 5%-10%

TOTAL STEPS

Preprocess data (convert columns to appropriate formats, handle missing values, etc.)

Conduct appropriate exploratory analysis to extract useful insights (whether directly useful for business or for eventual modelling/feature engineering).

Derive new features.

Reduce the number of variables using PCA.

Train a variety of models, tune model hyper parameters, etc. (handle class imbalance using appropriate techniques).

Evaluate the models using appropriate evaluation metrics. Note that is is more important to identify churners than the non-churners accurately - choose an appropriate evaluation metric which reflects this business goal.

In the end, choose a model based on some evaluation metric