

PROCESS to build a predictive model that can predict customer churn:

Step1: Business problem understanding

- Identifying the output variable.
- Importing the dataset.
- Understanding the variable.
- List of discrete and continuous variable.
- Created a test set, put it aside, and use it only to evaluate our Machine Learning models.

Step2: EDA. Pre-processing:

- Understanding the output variable.
- Encoding on output variable.
- Checking for null values
- Checked the outlier & skewness.
- Univariate analysis & bivariate analysis
- Understand the relation between inputs variables.
- Dropped the some input dependent columns to overcome multicollinearity problem.
- Encoding on input variables.
- Applied Scaling also but it decrease the accuracy and recall value scaling requires time to understand which column require scaling which not so not done the scaling. (Scaling requires time to do it in a correct way that's why not done)

Step3: feature selection:

- To select discrete columns used information gain.
- To select continuous columns used correlation matrix.
- Some rows are selected & dropped on the basis of EDA.
- Dropped the columns.

Step4: Balancing data set:

- Used both ways under sampling and oversampling it is performing well in under sampling. That's why performed under sampling.

Step5: Model Building:

- As per the provided data we can apply three algorithms Logistic Regression, Random Forest, Gradient Boosting. So I find best model among them only. But I can also apply more 15 algorithms to make a better model.
- Identified the best model on the basis of accuracy balanced precision, recall, f1score & auc-roc curve.
- Then build the final model after doing a hyper parameter tuning And build a final model with following results:
 1. Accuracy= 0.85
 2. Precision = 0.67
 3. Recall = 0.66
 4. F1_Score = 0.67

Step6: Model evaluation:

- Scores:
 1. Accuracy= 0.80
 2. Precision = 0.73
 3. Recall = 0.63
 4. F1_Score = 0.66
- Classification records
- Auc-roc curve