SYRIATEL TELCOM COMPANY PROJECT

INTRODUCTION

"Syriatel, a telecommunication company wants to Mitigate the loss of clients and boost its revenue.

I have been mandated the task to come up with the best solutions to reduce the loss of clients. "

DATA INFORMATION

Source: Kaggle

The dataset is made up of 3333 observations and a total of 21 columns. The target predicted is 'Churn' column.

There is an imbalance in class. There are 483 churns and 2850 no churns.

SMOTE or Tomek links may be needed.

EDA: Initial Observations

The variable, 'State', has different churn rates between each state. Due to this factor, I will retain it in my model.

EDA: Categorical Variable

1. Clients with 'international plan' tend to withdraw from the
2. service
3. Clients having more than '3 customer service calls' mostly
4. end up
5. churning.

EDA: Continuos Variable

Customers with a usage of over 280 min and customers who were charged $46 and above, caused a high churn.

Model Preparation

Replaced 'Yes' with ' 1 ' and 'No' with ' 0 ' for the columns 'international plan' and 'voice mail plan'

Dropped account length, area code, and phone number columns that do not have much significance.

The column 'State' was transformed into a dummy variable.

Total features : 66

Test size : 20%

Baseline Model

Recall score is 0 (diagram)

Logistic Regression.(diagram)

Tested Models

Random Forest

Logistic Regression

Decision Trees

XGBoost

Random Forest and XGBoost were chosen
for GridSearch

Parameter Tuning

XGBoost performed the best

Grid search Cross Validation was used to find optimal parameters

Best recommended parameters:

Final Model

Most Important features are:

  International plan

  Voice Mail Plan

  Customer Service Call

Recomendation And Future Execution

Recomendation:

Offer better services for international customers

Improve customer service experience. Prioritize on clients with frequent calls.

Future Execution

Investigate how the things you change (independent variables) affect the outcome you're measuring (dependent variable). Offer practical tips for digging deeper into these connections.

Eliminate features that have minimal impact on the model's performance. Then, retrain the models using the remaining, more relevant features.

To make the model more resilient (less sensitive to specific data points), we can increase data volume and address outliers that might disproportionately influence the model's learning.

THANK YOU