LEAD SCORING CASE STUDY

BY
SUBHRANIL SENGUPTA
STEVE THOMAS

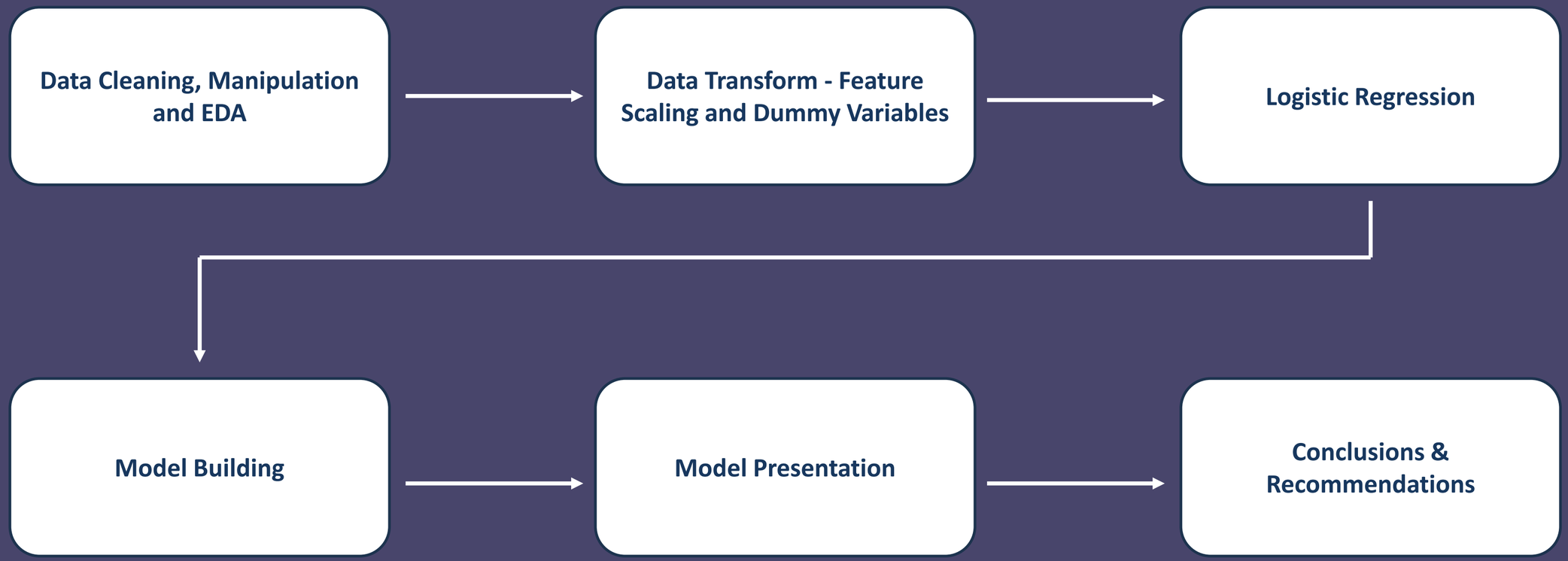
PROBLEM STATEMENT

X Education, an online education provider, faces a challenge with low lead conversion despite a steady influx of leads. To address this, the company aims to identify 'Hot Leads' – those with the highest potential for conversion. By implementing strategies such as lead scoring, data analysis, behavioral tracking, and personalized communication, X Education seeks to prioritize engagement with the most promising prospects. Additionally, optimizing the referral program and providing sales team training will enhance efficiency in lead management. Through continuous monitoring and refinement of these strategies, X Education aims to boost its lead conversion rate, drive revenue growth, and improve overall business performance.

BUSINESS OBJECTIVE

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%

APPROACH


Data Cleaning, Manipulation and EDA
Data Transform - Feature Scaling and Dummy Variables
Logistic Regression
Model Building
Model Presentation
Conclusions & Recommendations

DATA CLEANING

01

Dropping Null Values

Dropped columns with null values more than 45%

02

Data Skewness

Columns with skewed data were dropped

MARKET PROBLEM

01

Current Problem

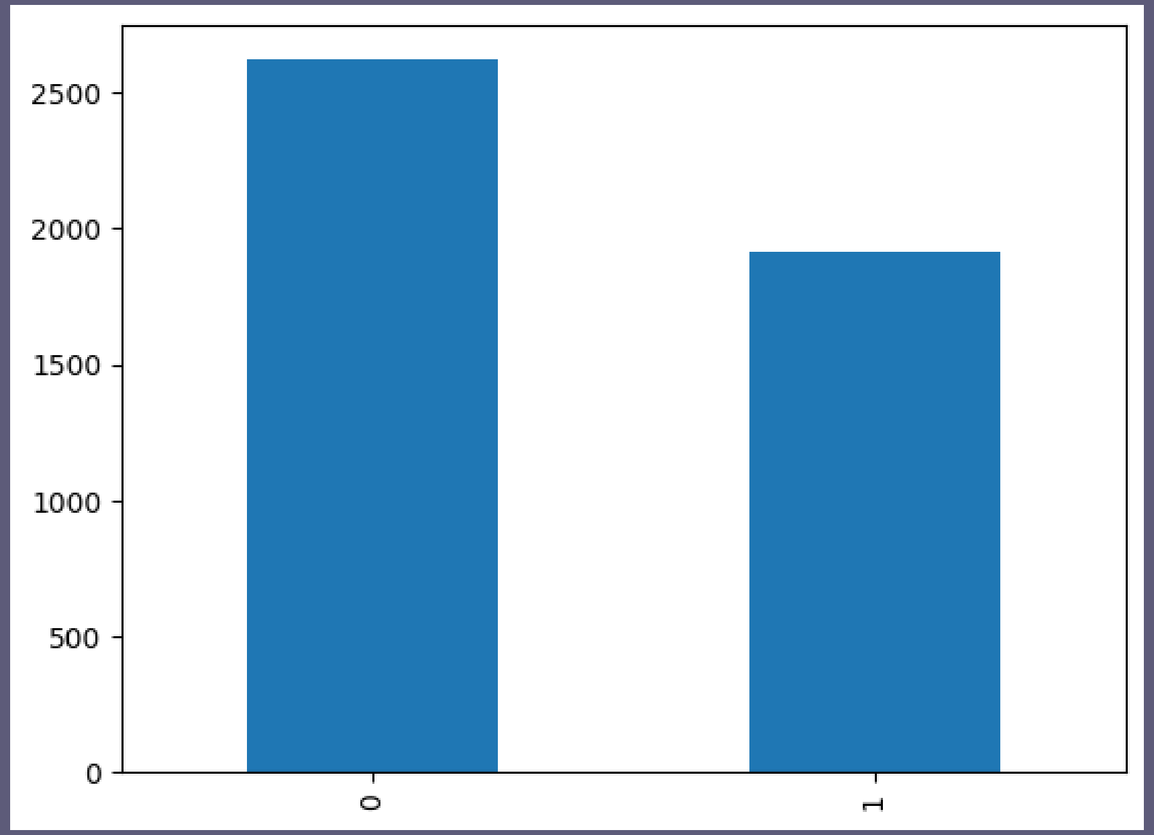
Write subtitle here
if you needed

02

Current Problem

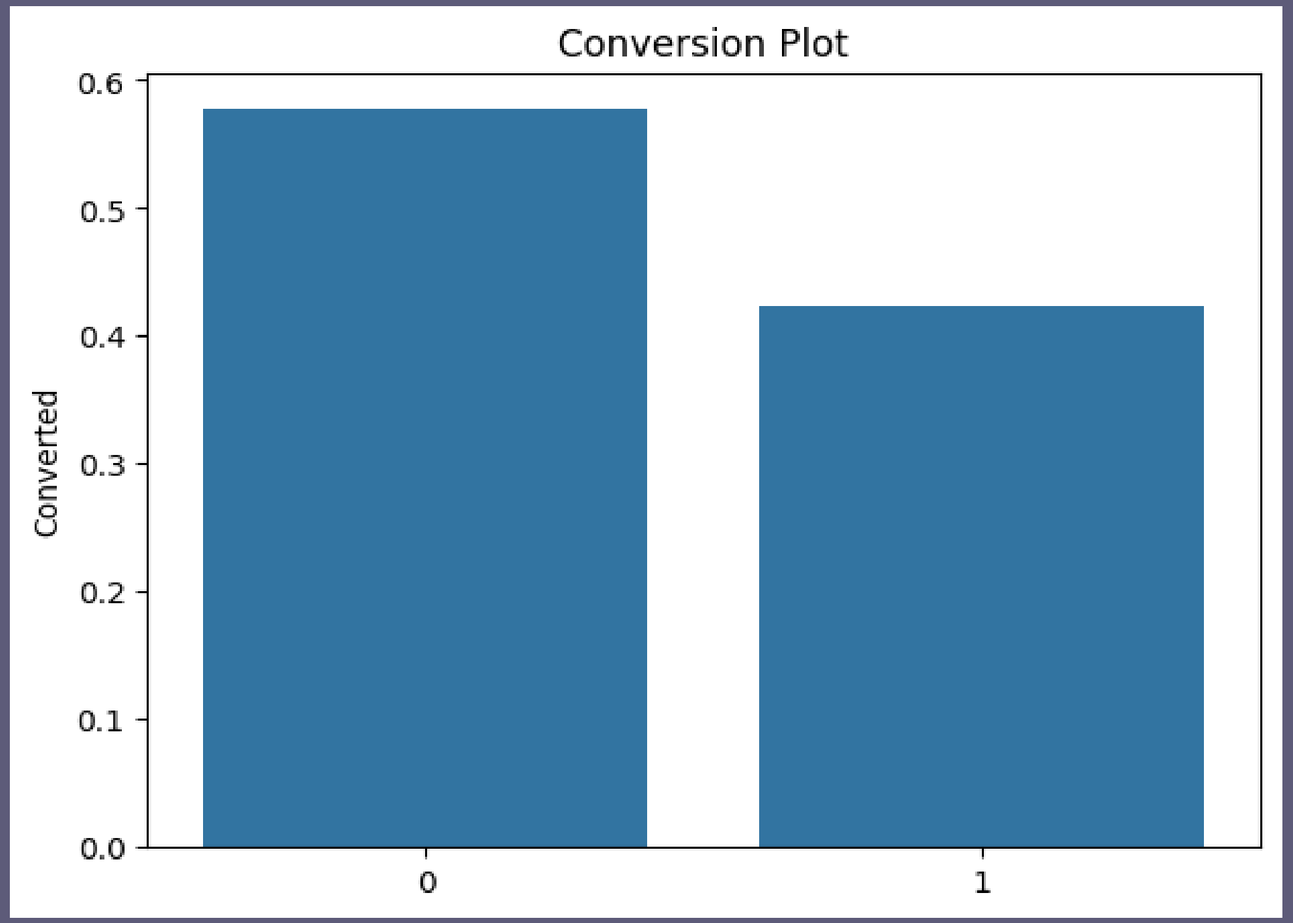
Write subtitle here
if you needed

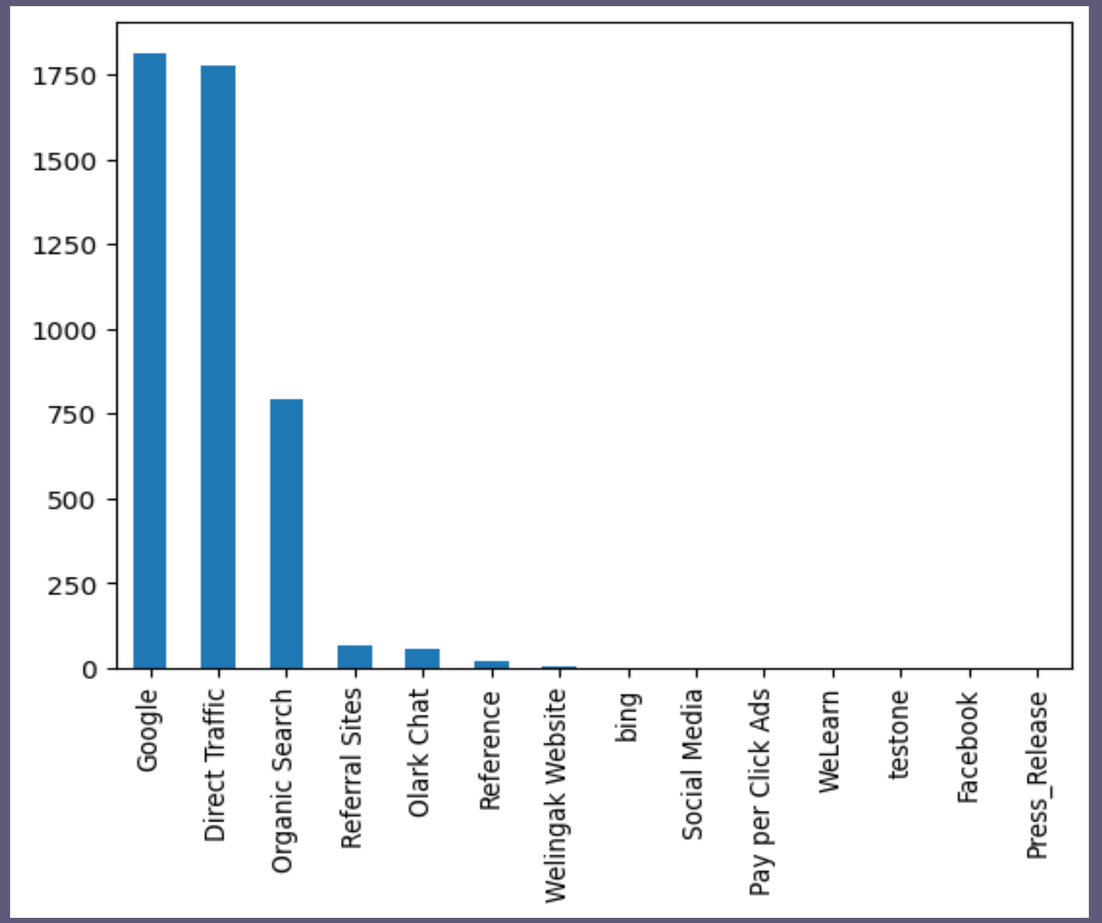
DATA IMBALANCE



Observation – ‘Converted’ :Successful lead conversion rate is just 42.3%. But, 57.7% of the Leads have not converted. So, the data is imbalanced.

VISUALIZATION OF UNIVARIATE ANALYSIS

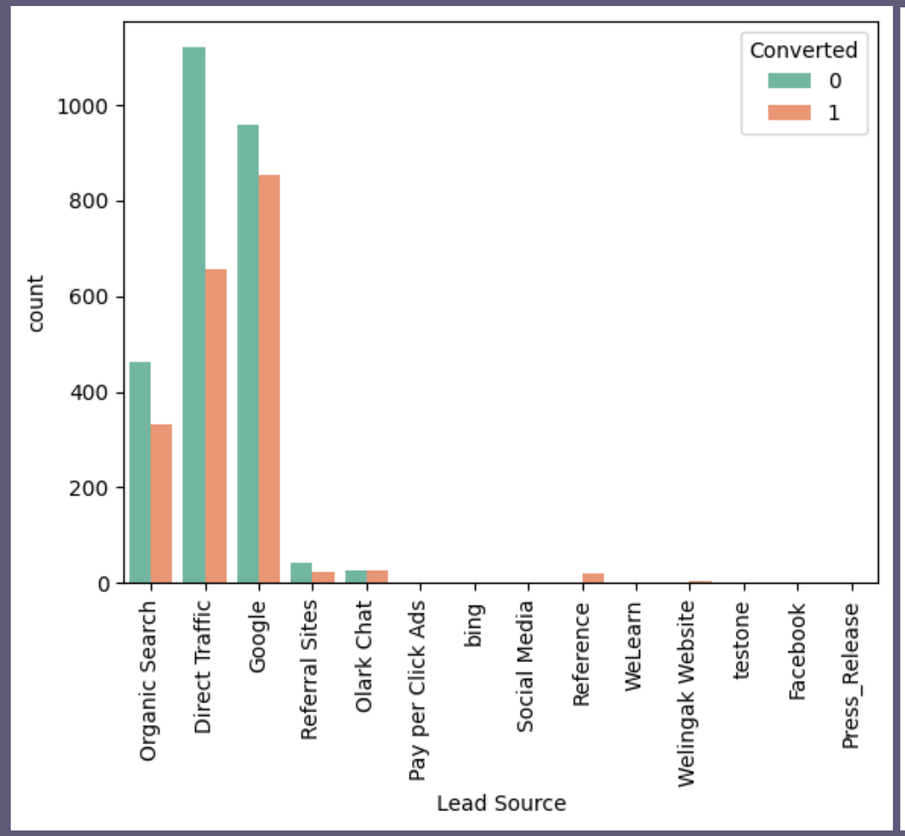




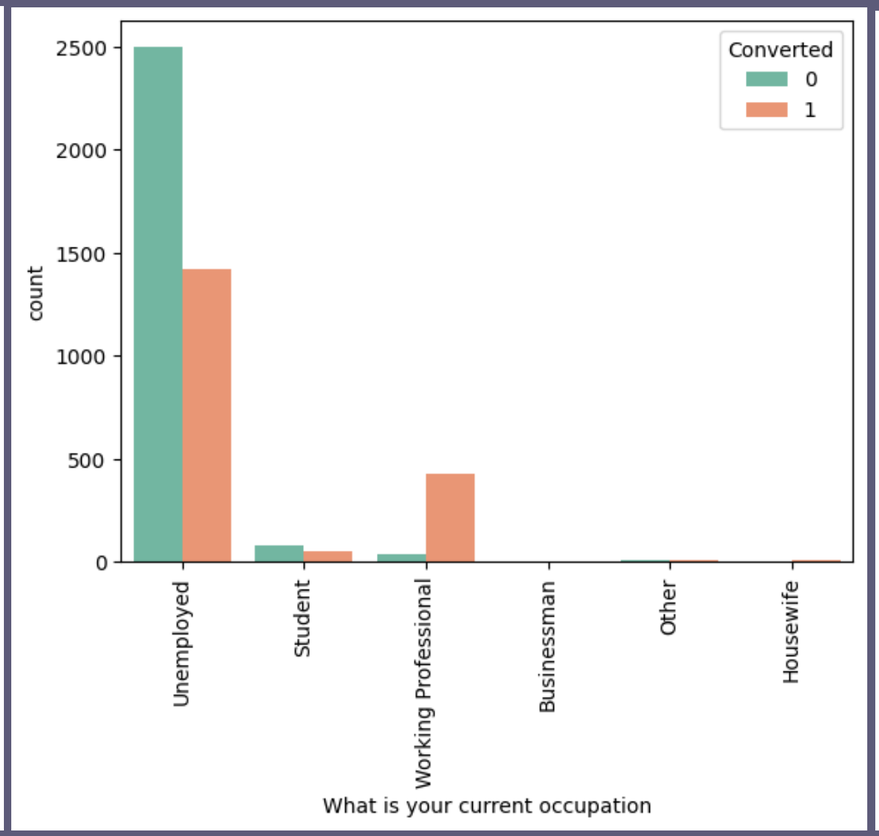
Observation – 'Converted' : Here we see that about 57% of the leads are not converted.

Observation: Most of the leads come from Google, Direct Traffic and Organic Search. X Education should put more emphasis on the mentioned sources while trying to increase the number of lists coming from Ads, Social Media and Press Release.

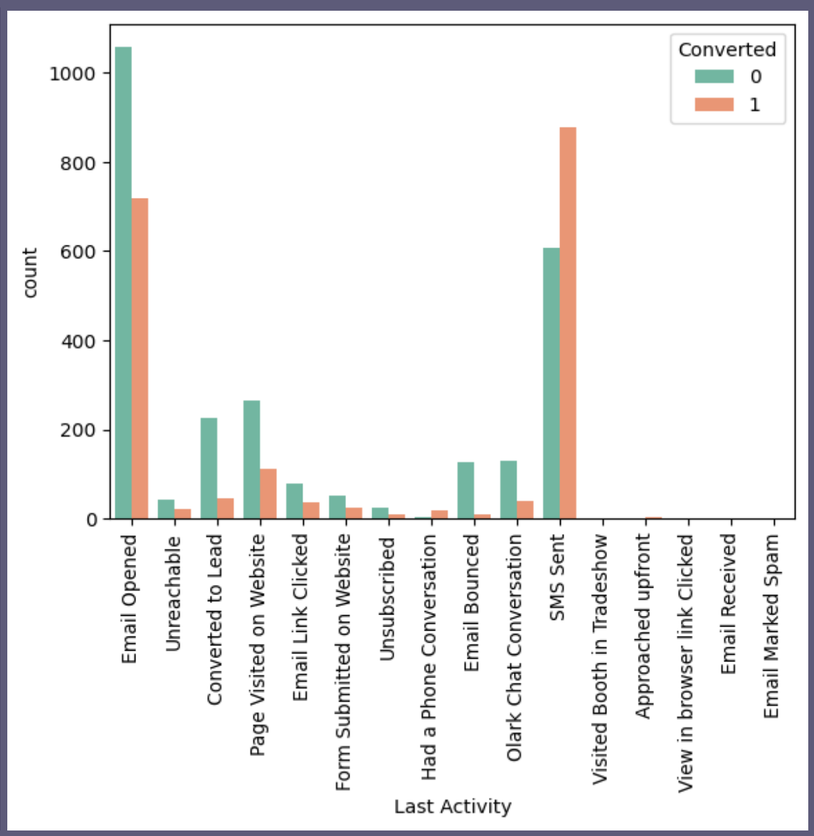
VISUALIZATION OF BIVARIATE ANALYSIS



Most leads come from Organic Search, Direct Traffic and Google. It is also important to notice that Reference has a very high conversion rate.



From the above plot, we can see that most leads are Unemployed. However, Working Professionals have the highest conversion rate.



It should be noted that "SMS Sent" has the highest lead conversion rate.

DATA TRANSFORM

01

Dummy Variables

Transformed categorical data

02

Scalling

Transformed numerical data uning StandardScaller

MODEL BUILDING

01

Basic Model

Built a model with all the features

02

RFE

Selected 25 features using RFE

MODEL BUILDING

03

VIF

Calculated VIF to further eliminate features

04

P-Value

Dropped 9 features due to high P-Value

MODEL BUILDING

05

Final Model

Final model is prepared with the featues

06

Metrics Evaluation

- Accuracy: 0.9622
- Recall: 0.9669

MODEL BUILDING

07

R.O.C.

ROC is plotted and the AUC is 99%

08

Cut-off

The cut-off is determined as 0.3

MODEL BUILDING

09

Prediction

The formulated model is fitted on the test data

10

Current Problem

- Accuracy: 0.9456
- Recall: 0.9571

MODEL FEATURES

FEATURE	COEFFICIENT
Tags_Closed by Horizzon	5.951381
Tags_Will revert after reading the email	4.258581
Country_Hong Kong	3.633706
Country_Germany	2.752979
Tags_Busy	1.285674
How did you hear about X Education_Other	1.030296
Total Time Spent on Website	0.979341
Last Activity_SMS Sent	0.854678
Lead Profile_Potential Lead	0.820431
Last Activity_Email Bounced	-1.628366
Tags_Interested in full time MBA	-2.261626
Lead Profile_Student of SomeSchool	-2.276199
Tags_Interested in other courses	-2.288052
Tags_Ringing	-2.400017
Tags_invalid number	-2.464825
Tags_Not doing further education	-2.720716
Tags_switched off	-2.827482

METRICS

MEASURES	VALUE
Accuracy	0.9456
Sensitivity	0.9572
Precision	0.9194
Recall	0.9572

RECOMMENDATION

- The evaluation metrics are pretty close to each other so it indicates that the model is performing consistently across different evaluation metrics in both test and train datasets.
- The model achieved a sensitivity of 96.22% in the train set and 94.56% in the test set, using a cut-off value of 0.3
- Sensitivity in this case indicates how many leads the model identifies correctly out of all potential leads that are converting
- The CEO of X Education had set a target sensitivity of around 80%
- The model also achieved an accuracy of 94.56% and Recall of 95.72% which is in line with the study's objectives

THANK YOU