Lead Scoring Case Study

This presentation outlines the work done by a group consisting of Namith Prakash, Namrata Ahirrao, and Pranali Prakash Naik to address a lead scoring problem for the online education company X Education.

Problem Statement

1

Low Conversion Rate

X Education sells online courses to industry professionals, but its conversion rate is quite low. Out of 100 leads acquired in a day, only about 30 are converted.

2

Need to Identify Hot Leads

To improve efficiency, the company wants to identify the most promising leads, known as 'Hot Leads', so the sales team can focus on these potential customers.

3

Business Objective

X Education wants to build a model that can identify these hot leads and deploy it for future use to increase their conversion rate.

Solution Methodology

Data Cleaning

The team performed various data cleaning steps, including handling duplicate data, missing values, outliers, and dropping columns with insufficient variance.

Data Analysis

They conducted both univariate and bivariate analysis to understand the data and identify patterns and relationships between variables.

Modeling

The team built a logistic regression model, performed feature selection using Recursive Feature Elimination (RFE), and validated the model to achieve an overall accuracy of 81%.

Data Manipulation

Initial Data

The initial dataset had 37 rows and 9240 columns.

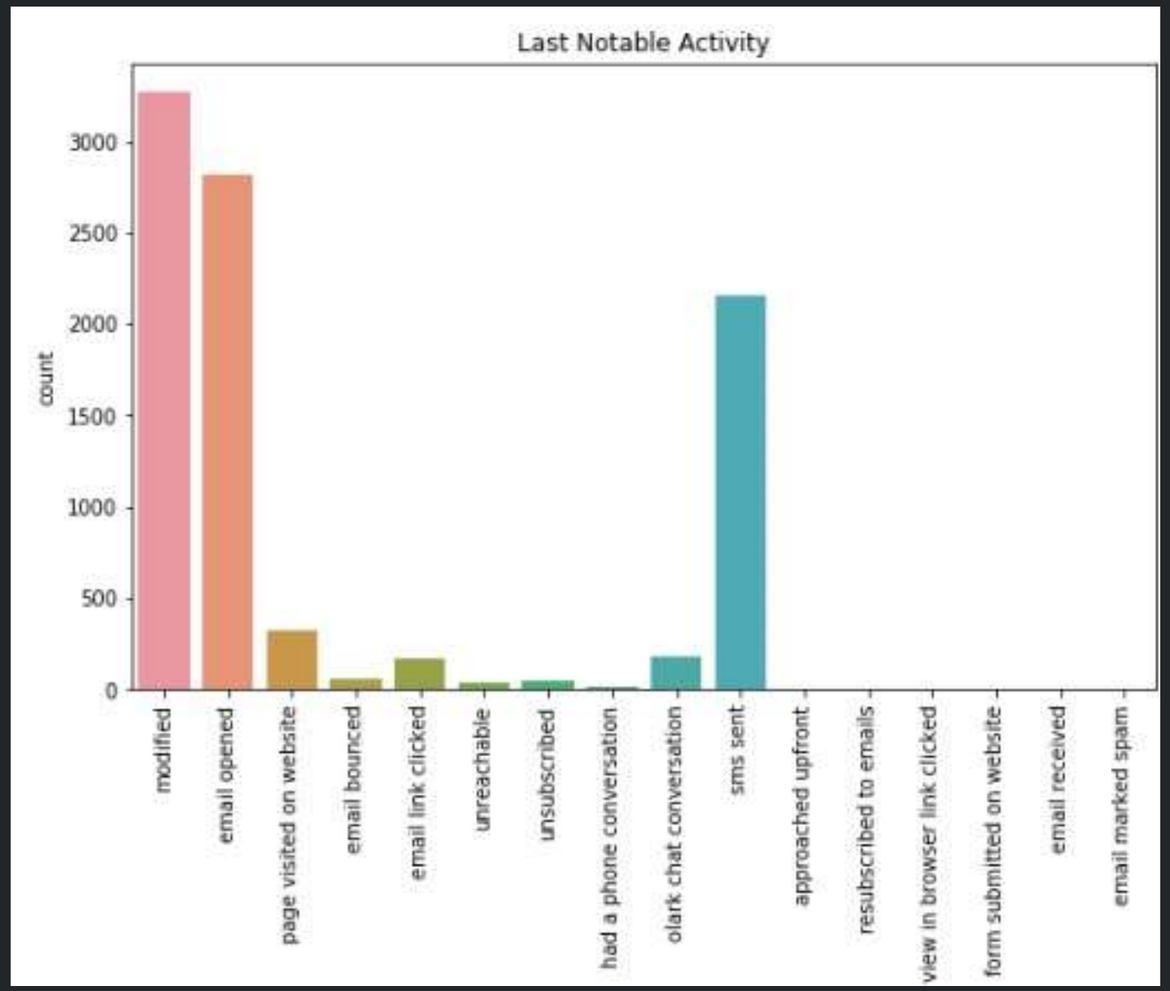
Feature Selection

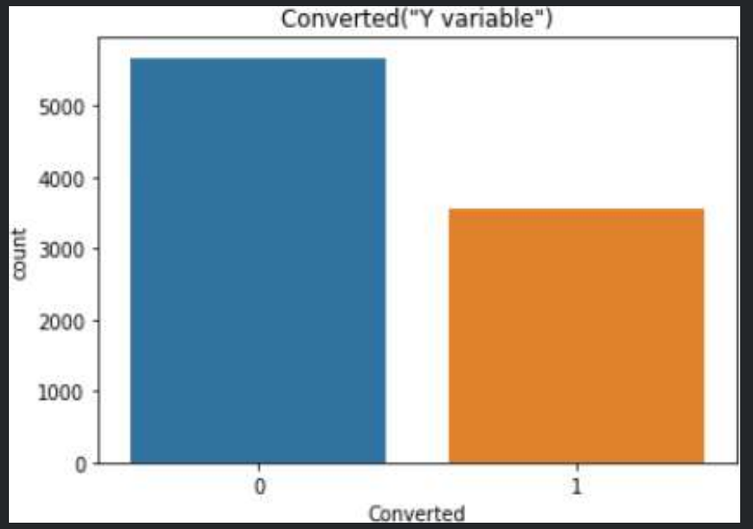
The team dropped single-value features, unnecessary columns, and features with more than 35% missing values to arrive at a final dataset of 8792 rows and 43 columns.

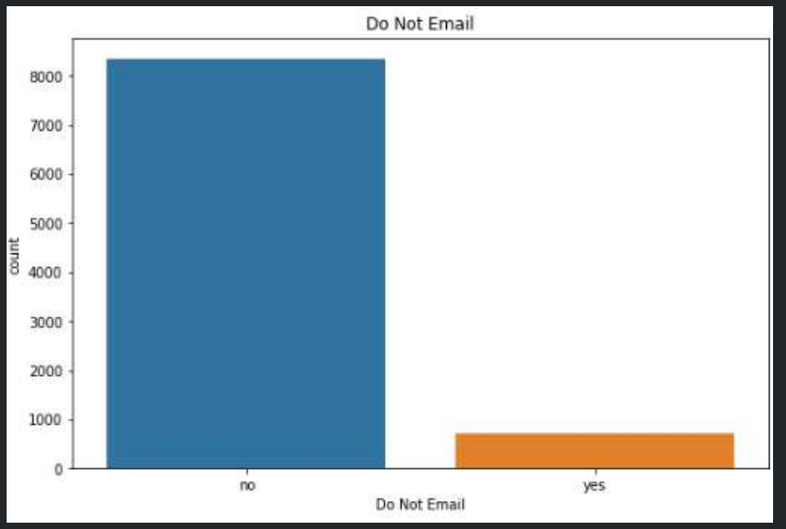
Data Conversion

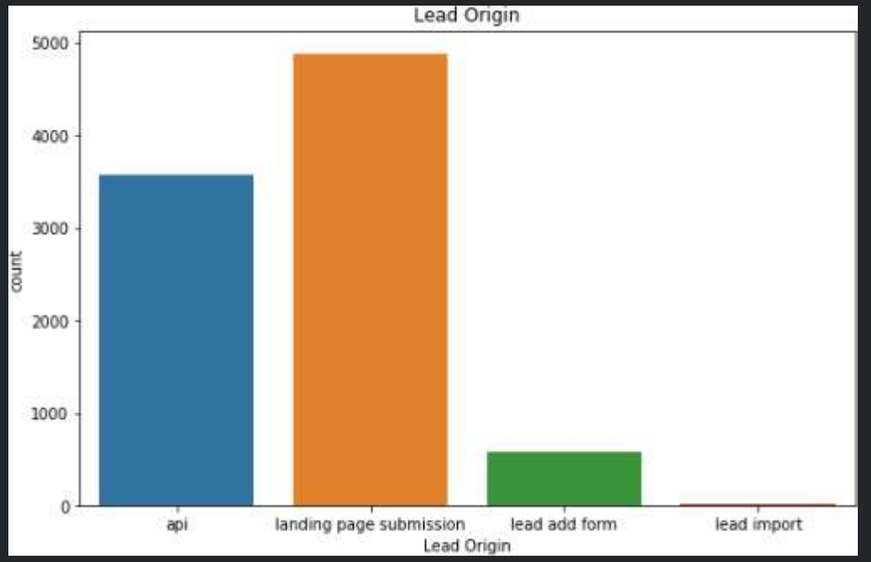
They normalized the numerical variables, created dummy variables for categorical variables, and split the data into training and testing sets.

Exploratory Data Analysis

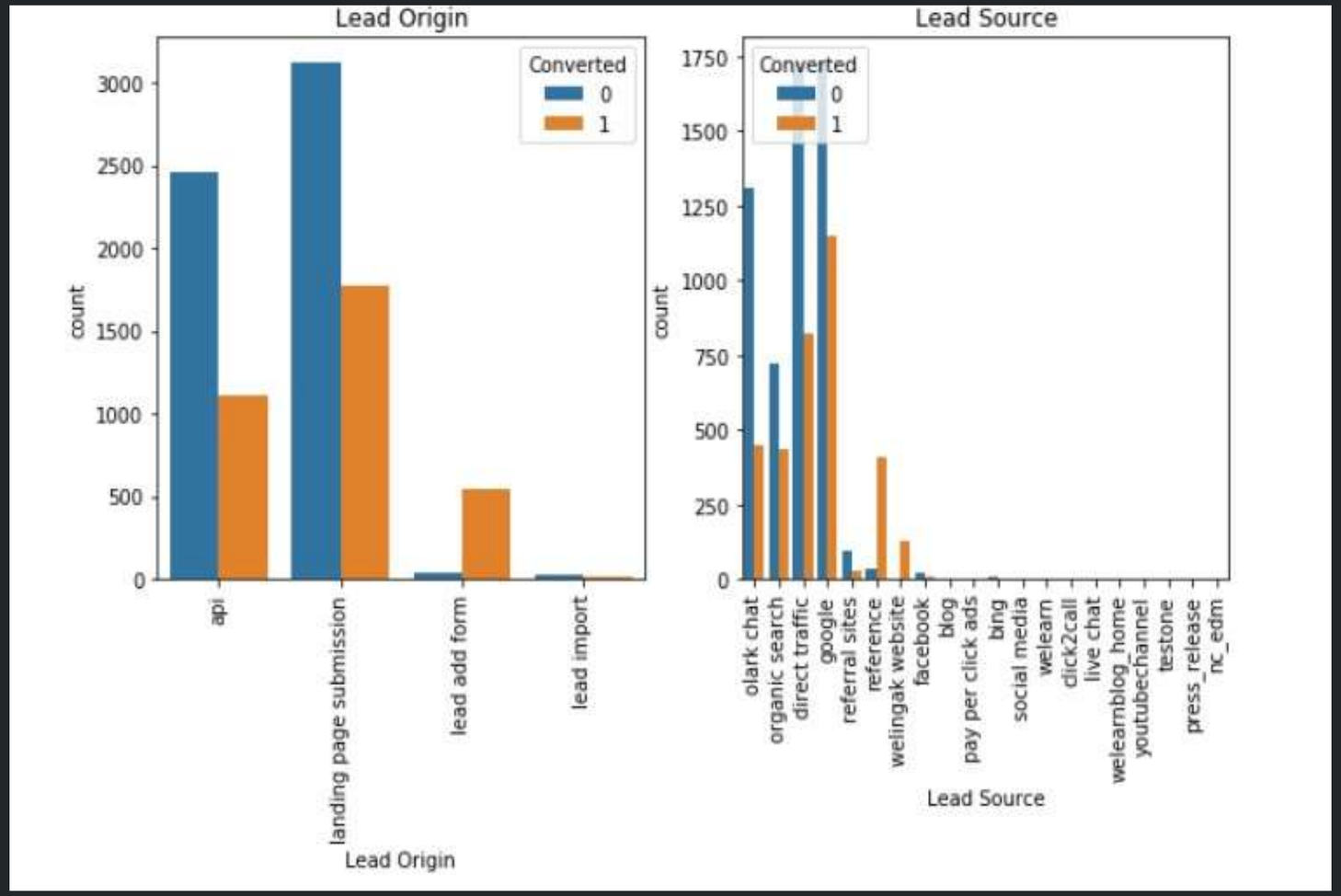


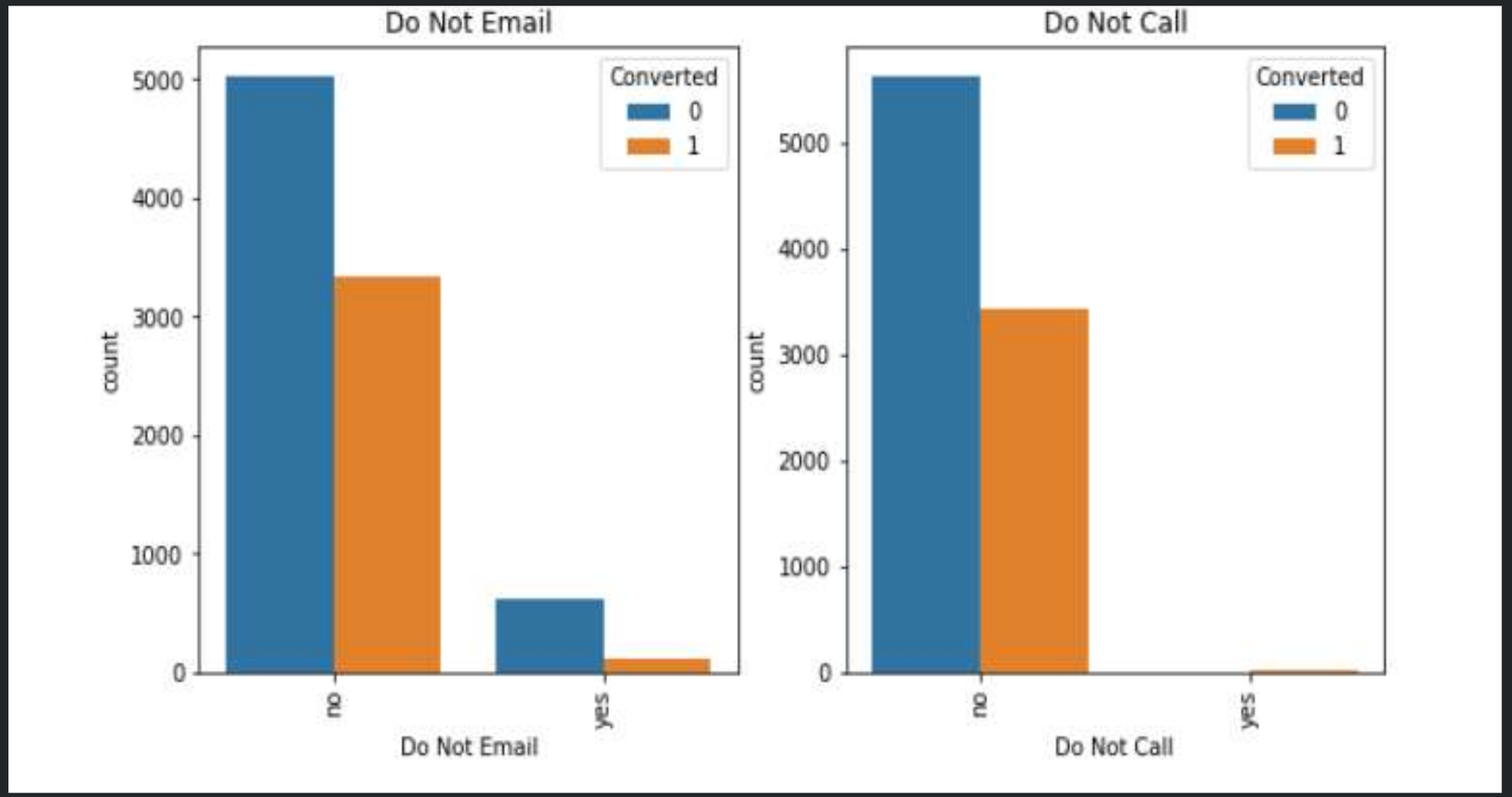

Converted("Y variable")
count
5000
4000
3000
2000
1000
0
0
1
Converted

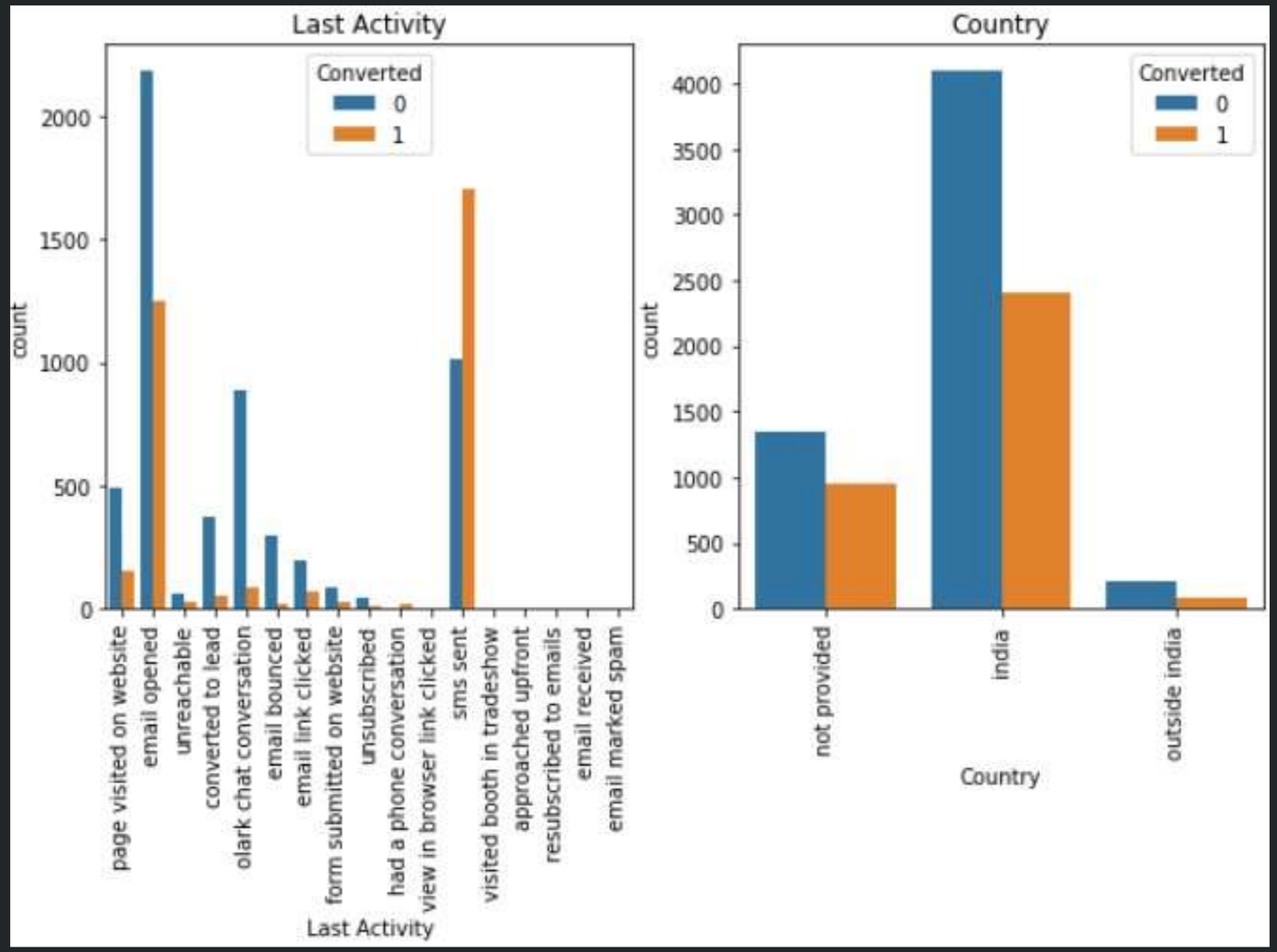

Do Not Email
count
8000
7000
6000
5000
4000
3000
2000
1000
0
no
yes
Do Not Email


Lead Origin
count
5000
4000
3000
2000
1000
0
api
landing page submission
lead add form
lead import
Lead Origin

Categorical Variable Relation




Do Not Email
Converted
0
1
5000
4000
3000
2000
1000
0
count
no
yes
Do Not Email
Do Not Call
Converted
0
1
5000
4000
3000
2000
1000
0
count
no
yes
Do Not Call


Last Activity
Converted
0
1
count
2000
1500
1000
500
0
page visited on website
email opened
unreachable
converted to lead
olark chat conversation
email bounced
email link clicked
form submitted on website
unsubscribed
had a phone conversation
view in browser link clicked
sms sent
visited booth in tradeshow
approached upfront
resubscribed to emails
email received
email marked spam
Last Activity
Country
Converted
0
1
count
4000
3500
3000
2500
2000
1500
1000
500
0
not provided
india
outside india
Country

Model Building

1

Feature Selection

The team used Recursive Feature Elimination (RFE) to select the top 15 variables for the model.

2

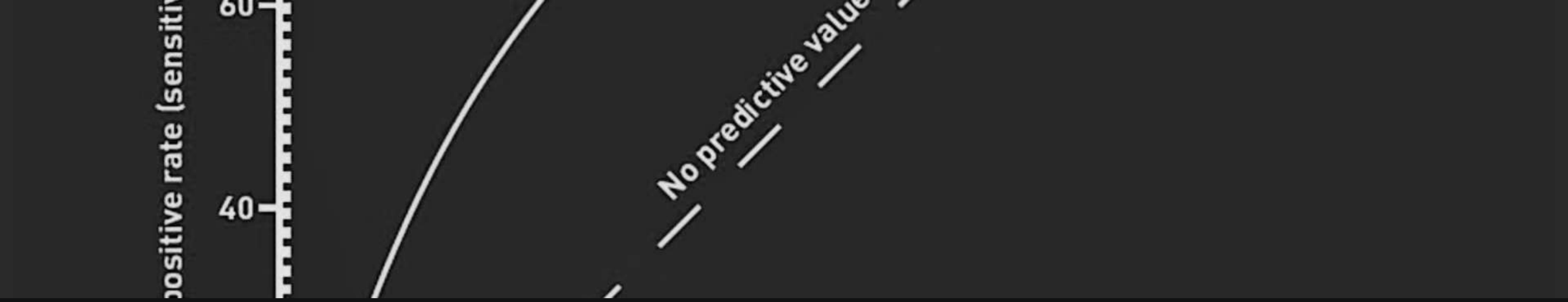
Model Training

They built the logistic regression model, removing variables with high p-values and VIF values.

3

Model Validation

The final model achieved an overall accuracy of 81% on the test dataset.



ROC Curve

Optimal Cut-off

The team determined the optimal cut-off probability to be 0.35, where sensitivity and specificity are balanced.

Key Factors for Potential Buyers

Time Spent on Website

Total time spent on the website is the most important factor for potential buyers.

Total Visits

The total number of visits to the website is also a key factor.

Lead Source

The most important lead sources are Google, direct traffic, organic search, and the Welingak website.

Last Activity

The last activity, such as SMS or Olark chat conversation, is another important factor.

Data Conversion

Numerical Variables are Normalised

Dummy Variables are created for object type variables

Total Rows for Analysis: 8792

Total Columns for Analysis: 43

Conclusion

1 Key Factors

The most important factors for potential buyers are total time spent on the website, total number of visits, lead source, and last activity.

2 Recommendations

By focusing on these key factors, X Education can significantly increase the chances of converting potential buyers into paying customers.