

Summary

This case study aims at building a Logistic regression model for X education company to predict potential leads and thus increase the Lead conversion rate from 30% to 80%.

Steps Followed are:

1. EDA:

- Dropped columns with more than 45% missing values.
- Replaced the NaN values with 'not provided' for categorical variable.
- Outlier treatment by eliminating top and bottom 1% values.
- Univariate, Bivariate and Multivariate analysis to gain useful insights.

2. Train-Test split & Scaling :

- Train-test split done at 70-30 %
- Min-max scaling for numerical variables ['TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website']

3. Model Building

- RFE was used for feature selection and top 15 relevant features are selected.
- Based on VIF and p-values, one by one variables are eliminated which are insignificant or multicollinear.
- A confusion matrix was created, and overall accuracy was checked which came out to be 81.04%.

4. Model Evaluation

- As per Sensitivity-Specificity view, the optimal cut off value is **0.36** &
- As per Precision – Recall view, the optimal cut off value is **0.44**

- **Sensitivity – Specificity**

Below model metrics are obtained using optimal cutoff of 0.36:

A) On Train Set:

- Accuracy = 80.64 %
- Sensitivity = 81.03 %
- Specificity = 80.40 %

B) On Test Set:

- Accuracy = 80.42 %
- Sensitivity = 80.70 %
- Specificity = 80.26 %

- **Precision – Recall:**

Below model metrics are obtained using optimal cutoff of 0.44:

A) On Train Set:

- Accuracy = 81.04 %
- Precision = 73.90 %
- Recall = 77.03 %

B) On Test Set:

- Accuracy = 79.05 %
- Precision = 73.36 %
- Recall = 70.12 %

CONCLUSION:

INSIGHTS FROM EDA:

- Target variable has Data imbalance. 62% peoples are not converted while 38% people are converted.
- Most of the lead sources are either from Google or they are Direct Traffic. While leads from Google and Organic Search have high conversion rate compared to others.
- People who are Unemployed have shown most interest in the course. While working professional people have shown a greater number of conversions followed by unemployed people. X-education should target Working professional people more for higher lead conversion.
- People who have expectation of “Better Career Prospects” have largely shown interest in this course with almost 50% conversion. X-

education should focus on providing more career opportunities along with course to get more lead conversions.

- People who have sent the SMS to X-education have shown huge conversion. X-education should focus on LEAD who have contacted X-Education through similar SMS.

TOP VARIABLE CONTRIBUTING TO CONVERSION:

- LEAD SOURCE:
 - Total Time Spent on Website
- Lead Origin:
 - Lead Add Form
- Lead source:
 - Welingak website
- Occupation:
 - Working professional

X education can use the above model on new leads to predict the probability of conversion for each LEAD. Also, it can use the model prediction to classify a LEAD as potential LEAD, to obtain 80% conversion rate.