

Lead Scoring Case Study

Summary

X Education faces a significant challenge with its lead conversion rate, currently hovering at a low 30%. To address this issue and meet the CEO's ambitious target of achieving an 80% lead conversion rate, the company has tasked us with developing a lead scoring model.

The goal is to increase the efficiency of lead prioritization and nurture leads effectively helping X Education move closer to its target of an 80% lead conversion rate and thereby ordering the company's bottom line.

Data Cleaning:

We began to eliminate the columns with a significant number of missing values, so we dropped the columns containing more than 45% of Null Value.

A new category named as 'Others' was created to make sense of the imputed values of the most frequent category.

Columns that did not contribute substantial value to our analysis were also dropped.

Exploratory Data Analysis:

Outliers were treated and capped to 95% for certain columns.

EDA was performed with respect to the target variable. As a result, different patterns were found to increase the low conversion rate of the leads.

Some of the important columns that the EDA found meaningful were Current Occupation, Lead Origin, Lead Source, Last Activity, Time Spent on the Website, etc.,

The Final Model:

Before building the Regression Model the dummy variables were created. The Data was split into two sets, one is the train set and the other is called the test set. This split was done on a ratio of 70:30.

Then using RFE the top 15 features were selected in order to build the Logistic Regression Model upon it.

Then VIF was calculated using the same 15 features. The variables with the value of less than 5 were kept and others were dropped. In this case none of the 15 variables were greater than 5.

Using the ROC curve the optimum cut-off value was measured. Then using the optimum cut-off value the accuracy, precision, sensitivity and specificity of the final model were

calculated.

Conclusion and Recommendation:

The final model has predicted a 90.7% of conversion rate.

The Accuracy of the final model is 80.6%, the sensitivity is 90.7%, specificity is 74.6% and the Precision is 69%

The top three features were Tags_Lateral_Student, Tags_Interested in Next Batch and Tags_Closed by Horizon.

Leads spending more time on the website are more likely to get converted.

Leads from Welingak Website has a conversion rate of 99%

Concentrating more on leads who open the email as the chances of getting converted is higher.

Target the Working Professional leads as the conversion rate is very high .