

Lead Score Case Study



Group Members
Sarthak Kapoor
Sanjay Poi
Shubham Jhadav

Problem Statement

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Outcomes

- Help them to select the most promising leads, i.e. the leads that are most likely to convert into paying customers
- Build a logistic regression model to predict the same
- Deploy for future use

Model Building Steps:

- Data Cleaning & Manipulation
 - Removed columns with null values greater 30% or 3000 in values
 - Dropping categorical variable which insignificant to the model
 - Dropped few rows for some columns with null values
 - Post all the cleaning we were left with 13 columns
- Feature Scaling/dummy variable for categorical columns
- Split the data in train and test
- Scaling the data using MinMaxScaling for both test and train
- Check for multicollinearity
- Proceed with Logistic regression using RFE to select top 15 variable
- Proceed to drop using insignificant columns using VIF and Logistic regression by 4 iterations
- Evaluated the Model using Confusion Matrix, Sensitivity, Specificity and Accuracy
- Area under ROC
- Find optimal cut off using accuracy, sensitivity and specificity graphs. Final value of 0.43
 - Accuracy 79.26
 - Sensitivity 79.71
 - Specificity 78.90
- Predict based on Test data
 - Accuracy 76.26
 - Sensitivity 83.9
 - Specificity 69.17
- Evaluation by Precision and Recall
 - Final accuracy 76.56
 - Precision 71.9
 - Recall 83.6

Conclusion

- Most important aspects of the platform which can be important for business to assess if the lead will get converted
 - Total visit & Total time spent on Website
 - Occupation
 - Housewife
 - Student
 - Unemployed
 - Working professional
 - When the last activity was
 - SMS
 - Olark chat conversation