

Lead Scoring Case Study Summary

The lead Score Analysis was conducted to help X Education in identification of more industry professionals to purchase their courses.. The basic data infers about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate.

Business Goal:

- Build Logistics Regression Model to predict the Potential lead
- The target lead conversion rate to be around 80%.
- A lead score between 0 and 100 to be assigned to each of the leads, which the company can use to target potential lead.

Approach taken:

Data Cleaning:

- 1• First step to clean the dataset we choose to remove the redundant variables/features.
- 2• Dropped the high percentage of Null values more than 40%. .
- 3• Identified the Highly skewed columns and dropped them.
- 4 • Treated the missing values by imputing the favourable aggregate function like (Mean, Median, and Mode).

Data Analysis

- We identified outliers in number of columns such as “Page_Views_Per_Visit” and “TotalVisits” .removed the Outliers.
- Identified variables that were having only single values and dropped them.

Data Preparation

- 1.Convert Binary Variables – Yes / No to 1 / 0
- 2.Create Dummy Variables – for Categorical.
3. Removed repetitive columns.
- 4.Used Standard scalar to scale the data for Continuous variables.
- 5.The Split was done at 70% and 30% for train and test the data respectively.

Model Building

- Used RFE technique for feature selection with 15 variable as output and fined tuned it manually by checking VIF and p-Values.
- Dropped columns which had high p values/VIF.

Model Evaluation

- A confusion matrix was created.
- Using ROC Curve, trade off between Sensitivity and Specificity Parameters, established optimal cut-off value of 0.35.
- • Ran Prediction for Train and Test Data Set---With the Current cut off of 0.35

Prediction

- Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of around 80%.

Conclusion•

The variables that mattered the most in the potential buyers are

- The total time spend on the Website.
- When the lead source was: Welingak website
- When the last activity was:
 - a. Phone Conversation
 - b. Olark chat conversation
- When the Lead Origin is Lead Add Form

With all these Observations X Education will be easily able to identify potential leads and convert them into a successful buyer.