

Summary of Lead Scoring Case Study

Following Steps we performed:

1. Importing the required libraries

2. Reading and Understanding the Data: Reading the dataset "Leads.csv" and Understanding it as follows:-

- a) Shape of data set(No of rows, columns in the dataset)
- b) Data type of each column, distribution, mean and median for all numerical columns etc.
- c) Checking the missing values
- d) Checking the duplicate rows

3. Data cleaning:

- a) Replaced "Select" value with NaN.
- b) Calculation of missing values for each column and dropping Score and Activity variable.
- c) Dropping the columns who have highest percentage of missing values.
- d) Checking the unique category for each column.
- e) If columns with categories are highly skewed, such columns are removed. Combine different categories in the column and reduce the percentage of the "Other" category.
- f) Imputing the column with the lowest percentage of missing values.

4. Exploratory data analysis: In this Step Univariate and Bi-Variate analysis was done on both categorical and numerical variables.

5. Outlier Treatment: We form outlier values for TotalVisits and Page View Per Visit.

6. Data Preparation: Data Preprocessing is performed in this step, dummy variables are created, performed train test data split and scaled the numerical columns.

7. Model Building and Model Evaluation:

- a) We used RFE feature selection method to get the final list of columns. In between the most insignificant, highly correlated columns are dropped and at last we had some columns in our final model.
- b) We choose the cutoff probability as 0.2 from Accuracy, Sensitivity, Specificity curve and calculated
- c) lead score for all the leads. The sensitivity of model was around 85%, specificity of model was around 96% , accuracy of model was around 91% and the conversion rate increased.

8. Conclusion: From model, we can conclude following points:

- a) A customer/lead who fills out a form is a potential lead.
- b) You should focus primarily on working professionals.
- c) You should primarily focus on prospects whose last activity was SMS sent or Email opened.
- d) It's always good to focus on the customers who spend a lot of time on our site.
- e) It's better not to focus too much on the customers whose sent emails are returned.

- f) If the lead source is a referrer, it may not be a potential lead.
- g) If a lead doesn't fill specialization, they may not know what to study and may not be the right audience. Therefore, it is better not to focus too much on such cases.

9. Recommendations

- a) We recommend collecting data and running the model frequently to keep your potential leads up to date.
- b) It is believed that the best time to call a potential leads is within a few hours of their interest in the course.
- c) Along with phone calls, it's good to mail the leads also to keep them reminding as email is as powerful as cold calling.
- d) Reducing the number of phone attempts to 2-4 and using other media such as Google ads or email to reach out to your leads more frequently can save you a lot of time.
- e) Focusing on Hot Leads will increase the chances of obtaining more value to the business as the numbers of people we contact are less but the conversion rate is high.