

# Summary Of Lead Scoring Case Study

After we have completed the reading of business understanding of this case study and the building of a machine learning model for the same. We got the goals of this case study. We have went through following some steps:

## **1. Data Sourcing :**

Importing the required libraries

## **2. Data Reading & Understanding :**

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

Routine Data Check: No of rows, columns, data type of each column, distribution, mean and median for all numerical columns etc. b. Missing value analysis. c. Duplicate rows check.

## **3. Data Cleaning:**

The data cleaning process is quite important in this case study. The data cleansing phase affects the model's effectiveness and quality. So, it must be carefully observed.

- a. NAN is used in place of the "Select" value.
- b. Calculating the value of each column's missing data and removing the Score and Activity variables.
- d. Eliminating columns with a large number of missing values.
- d. Verifying that each column has a distinct category.
- e. Columns that are heavily biased towards one category will be removed. incorporating various columns' categories with lower % values into the "Others" group.
- f. Input the column with the lowest percentage of missing values.
- g. Verifying the number of rows retained after completing the aforementioned procedures.

## **4. EDA :**

Both categorical and numerical variables were subjected to univariate and bivariate analysis in EDA.

## **5. Outlier Treatment:**

We form soft capping of upper range outlier values for TotalVisits and Page View Per Visit.

## **6. Data Preparation :**

In this step, We performed Data Preprocessing, the dummy variables are created. Performed train test data split and scaled the numerical columns.

## **7. Data Modelling & Model Evaluation:**

A. At first, there were 35 columns. The final list of columns was then obtained using both RFE and manual feature selection techniques. The least important and highly associated columns are removed in between, leaving our final model with 14 columns.

b. We are aware of how much more logical and understandable the relationship between  $\ln(\text{odds})$  of 'y' and the feature variable 'X' is.

The equation is:

c.  $\ln(\text{odds}) = -1.0565 * \text{const} + 0.1944 * \text{TotalVisits} + 1.0574 * \text{Time Spent} - 0.3186 * \text{Free Copy} - 1.0199 * \text{Lead Origin\_Landing Page Submission} + 4.4017 * \text{Lead Origin\_Lead Add Form} + 1.2101 * \text{Lead Source\_Olark Chat} - 1.1764 * \text{Lead Source\_Reference} - 1.1921 * \text{Last Activity\_Email Bounced} + 0.8166 * \text{Last Activity\_Email Opened} - 0.6859 * \text{Last Activity\_Olark Chat Conversation} + 0.6463 * \text{Last Activity\_Others} - 1.9097 * \text{Last Activity\_SMS Sent} - 1.1380 * \text{Specialization\_Not Specified} + 2.6908 * \text{Current Occupation\_Working Professional}$

d. We chose the cutoff probability as 0.35 from Accuracy, Sensitivity, Specificity curve and calculated lead score for all the leads. The sensitivity of the model was around 80 % and the conversion rate increased from 38% to 68%.

## **8. Conclusion:**

We can draw the following conclusions from the model: Potential leads are those who fill out the form as customers or leads. The working class must be our main priority. The leads on whom an email or SMS was opened last must be our primary emphasis. Focusing on clients who have invested a lot of time on our website is always a good idea. It's advisable to pay least attention to the clients to whom they sent mail that was returned. If the lead source is a referral, the potential lead could not actually be that

person. The lead may not know what to study and is not the ideal person to target if the speciality was not filled out by the lead. Thus it's best to pay less attention to such cases. Run the model to update your knowledge of prospective leads. It's generally accepted that the optimal time to call prospective leads is shortly after they express interest in your courses. As email is just as effective as cold calling, it's a good idea to mail the leads in addition to making phone calls to remind them. It will save a lot of time to use other media, such as Google ads or emails, more frequently to stay in touch with leads and cut down on the number of call attempts to 2-4. Concentrating on Hot Leads will boost the likelihood of gaining more value for the company because we contact fewer people but they convert at a higher rate.