
Summary -Lead Scoring Case study

As per the understanding of the business problem, we got the objectives and goals of our case study. According to the data received, we have done the following steps to perform our analysis.

- 1) **Data Sourcing:** Imported the required libraries
- 2) **Data reading and understanding:** Uploading dataset using the command “pd.read_csv('Leads.csv’)” and understand the data set performing routine data check (shape of the data set, info, mean median statistics, data type etc), Missing value analysis, Duplicate rows check etc.
- 3) **EDA:** Data cleaning plays very crucial role here, as the quality of the model depends on the data cleaning step. Hence it must be followed thoroughly.
 - a) “Select” Value is replaced with NAN.
 - b) Calculation of missing values
 - c) Dropping columns with 45% of missing values.
 - d) Checking unique value of columns and dropping them accordingly.
- 4) **Analysis of Categorical & Numerical Attributes:** compared each column through the plot distributions and dropped certain columns which are not necessary for the analysis.
- 5) **Outlier Treatment:** We performed outlier range values for Total Visits, Total Time Spent on Website, & Page Views Per Visit.
- 6) **Data preparation:** In this step, we had created dummy variables and performed train data test split and scaled the numerical columns.
- 7) **Modelling Building and Model evaluation:** When we got the data, we had 37 columns after using RFE methods and dropping highly correlated columns we had 15 columns in our final model.

We choose the cut-off probability as 0.35 from Accuracy, sensitivity, specificity curve and calculated lead score for all leads. The Sensitivity of the model was around 80%.
- 8) **Conclusion:** Overall our model seems to predict the Conversion Rate well and able to give the CEO confidence in making good calls based on this model, and we recommend collect data often and run the model and get updated with the potential leads thus we can concentrate more on hot leads which will help us to boost our business.

As per the model: Variables matter the most in the potential buyers are:

 - a) Total Visits (The customer who visits our websites & the total time spend on the Website.
 - b) Lead Origin Lead Add Form (The customer/leads who fills the form are the potential leads.)
 - c) Lead Source Direct Traffic, Lead Source Google, Lead Source_Welingak Website, Lead Source_Organic Search, Lead Source_Referral Sites, Lead Source_Welingak Website, Do Not Email_Yes, Last Activity_Email Bounced, Last Activity_Olark Chat Conversation