

Summary : Case Study – Lead Score

➤ Exploratory Data Analysis:

- ✚ It is Observed that the Lead Source graph "Google", "direct traffic", "Organic search", and "olark chat" have higher Counts. It is observed that most of the customer is from INDIA. Major conversion in the lead source from google.
- ✚ Most of the customers are unemployed, choose courses for "Better Career Prospects" and they are from "Mumbai".
- ✚ The total conversion rate of customers is 37.82%
- ✚ Leading Page Submission from Lead Origin are Having More Total Time Spent on Website, Total Visits and Page Views Per Visit compared to other Category Levels and Customers who took the Course spent more time on the website.

➤ Model Building and Evaluation:

- ✚ Final Model with 13 Features. Features like Total time spent on the website, Lead source_welingal Website and Lead origin_lead Add form are positively related to the response variable. Features like Do not Email, Lead_profile Student of some school, Last Notable Activiti_Olark Chat Conversation, etc. are Negatively Correlated to Response Variable.
- ✚ It is Observed that Sensitivity and Specificity intersected at a probability of 0.4 and Precision and Recall also tell that a probability of 0.4 is the Optimal cut-off for the model.
- ✚ The area under the Curve (AUC) is 0.89 which is Obtained from Receiver Operating Characteristic Curve.

✚ Confusion Matrix and metrics on Training Data:

Actual/Predicted	Not Converted	Converted
Not Converted	3535	392
Converted	770	1597

Metrics:

Metric	Value
Accuracy	0.81
Sensitivity or Precision or True Positive rate	0.67
Specificity	0.9
False Positive rate	0.09
Positive Predicted value	0.8
Negative Predicted value	0.82

Confusion Matrix and metrics on Test Data:

Actual/Predicted	Not Converted	Converted
Not Converted	1650	14
Converted	776	258

Metrics:

Metric	Value
Accuracy	0.7
Sensitivity or Precision or True Positive rate	0.24
Specificity	0.99