

SUMMARY

The analysis is being done to for company X Education and to obtain ways to convert potential users into joining their course. The data available is about the information of the potential customer, the site they visit and their conversion rate.

Steps:

Data Understanding: To Get an overview of the data.

Cleaning the data: Handling the null values. The columns with greater than 45% null values in the dataset were dropped. Replaced the NaN values in the dataset.

EDA: EDA was done on the dataset. Countplot was plotted for the variables along with the converted leads to visualise them. Boxplot was plotted to check the outliers on the columns.

Dummy Variables: The dummy variables were created for Categorical columns.

Train-Test Split: The split was done with the ration of 70:30, The Train data of 70% and Test Data of 30%.

Model Building: RFE was used for feature selection to attain the Top 20 variables. Then the rest of the variables were manually dropped based on their VIF values and the P-value ($VIF < 5$ and $P\text{-value} < 0.05$).

Model Evaluation: To Evaluate the model a confusion matrix was created. Then, ROC curve was used to find out the optimum cutoff value.

- **Sensitivity – Specificity**
 - **On the Training Data:**

Accuracy – 81.24%
Sensitivity - 79.08%
Specificity - 82.55%.

- **On the Test Data:**

Accuracy - 80.42%
Sensitivity - 80.03%
Specificity - 80.67%.

- **Precision – Recall:**
 - **On the Training Data:**
 - Accuracy - 81.55%
 - Precision - 74.57%
 - Recall - 77.31%
 - **On the Test Data:**
 - Accuracy - 80.57%
 - Precision - 73.34%
 - Recall - 77.32%

FROM THE MODEL THESE ARE THE TOP VARIABLES WHICH ARE CONTRIBUTING TO CONVERSION:

- **TOTAL TIME SPENT ON WEBSITE**
- **WHAT IS YOUR CURRENT OCCUPATION_WORKING PROFESSIONAL**
- **LEAD ORIGIN_LEAD ADD FORM**

Lead Source:

- **Google**
- **Direct Traffic**
- **Organic Search**
- **Referral Sites**
- **Welingak Website**

Lead Origin:

- **Lead Add Form**

Last Activity:

- **Olark Chat Conversation**
- **Email**
- **Page Visited on Website**

The company by focusing on these areas can convert most of the potential users into joining their course.