

Summary

Objective

The objective of the case study was to build a logistic model to assign a lead score between 0 and 100 to each of the leads, which can be used by the company to target potential leads with a high probability of conversion (>80%).

Methodology

The following steps were performed:

Data Understanding, Cleaning, and Preparation

- All variables in the dataset were analysed based on domain knowledge.
- Duplicate entries were checked and found to be absent.
- Exploratory data analysis (EDA) was performed, treating missing values and handling outliers for numerical features.
- The distribution of data was analysed using count plots, and correlation for numerical variables was checked using a heatmap.
- The overall conversion rate was found to be 38%.

Train-Test Split and Scaling

- The dataset was split into train and test sets with a 70:30 ratio and a random_state of 100.
- Numerical features were standardized.

Model Building

- The top 15 most important variables were selected using recursive feature elimination (RFE).
- A logistic regression model was built using these variables, and the model was analyzed with respect to parameters such as variance inflation factor (VIF) and p value.

- Variables were dropped if the p-value was greater than 0.05 or VIF was greater than 5.
- The optimal model was obtained with $VIF < 5$ and p-values < 0.05 for all variables.

Model Evaluation

- Various model metrics were evaluated, including accuracy, sensitivity, and specificity.
- A receiver operating characteristic (ROC) curve was plotted to determine the threshold, which was found to be 0.3.
- Other metrics such as precision, recall, and F1_score were also examined.
- The results obtained on the train set for the ideal cutoff of 0.3 were:
 - Accuracy – 81.0 %
 - Sensitivity – 81.7 %
 - Specificity – 80.6 %
- The model was then evaluated on the test set, and the results were similar to those obtained on the train set:
 - Accuracy – 80.4 %
 - Sensitivity – 80.4 %
 - Specificity – 80.5 %

Key Factors Impacting Lead Conversion Rates

The following features were found to have the greatest impact on lead conversion rates:

- The company should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference" as these are more likely to get converted.
- The company should make calls to the leads who are the "working professionals" as they are more likely to get converted.
- The company should make calls to the leads who spent "more time on the websites" as these are more likely to get converted.
- The company should make calls to the leads coming from the lead sources "Olark Chat" as these are more likely to get converted.
- The company should make calls to the leads whose last activity was SMS Sent as they are more likely to get converted.
- The company should not make calls to the leads whose last activity was "Olark Chat Conversation" as they are not likely to get converted.
- The company should not make calls to the leads whose lead origin is "Landing Page Submission" as they are not likely to get converted.
- The company should not make calls to the leads whose Specialization was "Others" as they are not likely to get converted.
- The company should not make calls to the leads who chose the option of "Do not Email" as "yes" as they are not likely to get converted.

•