

Leads Scoring Case Study

Summary

Steps how we have proceeded with our assignments are as follows:

1. Data Cleaning:

- a. First step to clean the dataset we removed columns having more than 40% null values.
- b. Then we choose to remove the redundant variables/features.
- c. After that we found that some columns are having label as 'Select' which means the customer has chosen not to answer this question. The ideal value to replace this label would be null value as the customer has not opted any option. Hence, we changed those labels from 'Select' to null values.
- d. For remaining missing values, we have imputed values with maximum number of occurrences for a column.
- e. For low null values columns we dropped the rows.

2. Data Transformation:

- a. Changed the multicategory labels into dummy variables.
- b. Checked the outliers and handled them.
- c. Removed all the redundant and repeated columns.

3. Data Preparation:

- a. Split the dataset into train and test dataset and scaled the dataset.
- b. After this, we plot a heatmap to check the correlations among the variables.
- c. Found some correlations and they were dropped.

4. Model Building:

- a. We created our model with RFE count 15 and removed the insignificant columns by comparing with p-values and VIF values..
- b. For our final model we checked the optimal probability cutoff and checking the accuracy, sensitivity and specificity.
- c. Convergence of accuracy, specificity and sensitivity gave us the optimal cutoff value.
- d. We checked the precision and recall with accuracy, sensitivity and specificity for our final model and the tradeoffs.
- e. Prediction made in test set and predicted value was recorded.
- f. We did model evaluation on the test set and found the score of accuracy and sensitivity for our final test model is in acceptable range.

5. Conclusion:

- i. Test set is having accuracy, recall and sensitivity in an acceptable range.
- ii. Top features for good conversion rate:
 - 1. **TotalVisit**
 - 2. **Total Time Spend On Website.**
 - 3. **Lead Origin_Lead Add Form**
 - 4. **Lead Source: Olark chat conversation, reference and wellingal website.**
 - 5. **When the last activity was Olark chat conversation and SMS sent.**
 - 6. **When their current occupation is as a Working Professional.**