

Summary

We tried to solve a problem that company X had with converting leads into potential users. We started performing key steps to understand data and clean it. We further built few models and tested them bases on certain factors. Our goal was to assign a lead score to the leads which will help the company to increase lead conversion. The steps we followed were:

1. EDA:

- A quick check was done for nulls, and we dropped columns with more than 40% missing values.
- We also saw that null rows would cost us a lot of data and were important columns. Instead, we replaced NaN values .
- We also worked on numeric variables, outliers and dummy variables.

2. Train-Test split & Scaling:

- The split was done at 70% and 30% for train and test data respectively.
- We will do min-max scaling on the variables ['TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website']

3. Model Building:

- RFE was used for feature selection.
- An RFE was then conducted to obtain the top 15 relevant variables.
- Later, the rest of the variables were removed manually depending on the VIF values and p-value.
- A confusion matrix was created, and the overall accuracy was checked which came out to be 79%.

4. Model Evaluation:

- Evaluation Metrics for the train Dataset: -

 Accuracy :0.79

 Sensitivity:0.78

 Specificity:0.80

 Precision: 0.72

 Recall: 0.78

- Evaluation Metrics for the test Dataset: -

✚ Accuracy :0.78

✚ Sensitivity:0.78

✚ Specificity:0.80

✚ Precision: 0.70

✚ Recall: 0.76

5. Conclusion:

- Lead Origin_Lead Add Form: Leads who participate in the "Lead Add Form" have a higher conversion rate, so companies can focus on getting more leads because there is a higher chance of change.
- What is your current occupation_Working Professional: Managers with "Employee Professional" job have more flexibility, companies should focus on working professionals and try to get more leads.
- Total Time Spent on Website: Administrators who spend a lot of time on the site can do it for us.

The Model seems to predict the Conversion Rate very well and we should be able to give the Company confidence in making good calls based on this model.