



LEAD SCORING CASE STUDY SUMMARY

Abhishek Jain



This analysis has been done for a company called X Education which wants to find better methods to sell online courses to professionals. We are required to understand and validate the given data to help the company target the correct set of people and increase its lead conversion rate. The steps followed for the analysis are:

EDA

- Replaced the 'Select' values in columns.
- Dropped columns with more than 25% missing values.
- Replaced the NaN values with average values in the column.
- Also analyzed data on numerical variable, outliers and dummy variables.

Training and Test data

- The split for train and test data was done at 75% and 25% respectively.
- Min-max scaling was done on the variables 'TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website'.

Model Building

- RFE was used for feature selection and used to attain top 25 relevant variables.
- Next, variables were removed manually as per the VIF and p values.
- Overall accuracy was checked which came out to be around 80%.

Model Evaluation

- The optimum cut off value was determined using ROC curve and the area under ROC curve was **0.86**.
- We found that optimum cutoff was **0.36**.
- Overall accuracy of Training data came out to be around **80%**.
- Overall accuracy of Test data also came out to be around **80%**.

Conclusion

The top three variables for probability of lead conversion as per our model are:

- Total Time Spent on Website
- Last Activity_SMS Sent
- Last Activity_Olark Chat Conversation

The top three variables in our model, that should be focused for increasing the probability most are:

- Last Activity_SMS Sent (positively impacting)
- Lead Source_Google (negatively impacting)
- Lead Source_Email Opened (negatively impacting)