

Summary Report

An education company named X Education sells online courses to industry professionals. Based on the data set containing lead data, X education needs help in identifying the driving factors that are indicators of a lead getting converted to an actual paying customer.

Methodology and Observations

After data preparation, features were selected based on RFE before building a prediction model on these variables to identify the probability of lead conversion. First step was to prepare the data by handling null values based on the percentage of null values in a particular column. This also involved standardization of values as the data was not uniformly segregated across the data set. Next step is to split the data set into test and train data sets for training and once the model is optimized the same can be used in the test data set. Fit transform was run on the continuous variables before calculating lead conversion rate of the data set which came out to be 48%. After correlation analysis features with more than 0.8 correlation was discarded as they were highly correlated with each other. Next, an iterative feature selection was run on the train data set to identify the optimal number of features with the most accuracy and this number turned out to be 33. After various rounds of VIF and P value calculation, we removed a few more variables as they either had a high VIF value indicating multicollinearity or a high P value with a low VIF indicating the variable is not statistically significant. Next step is to assess the model for any class imbalances using confusion matrix and ROC curve and both returned acceptable results. The precision and recall of the train data set came out to 0.79 and 0.73 respectively. Finally, we went on to apply the model on the test data set and the accuracy, sensitivity and specificity of the model are 0.76, 0.73, 0.80 respectively.

Conclusion

Based on the data provided, "Lead Profile" specified as "Potential Lead", "Total Time Spent on Website" and "Last Activity" specified as "Had a Phone Conversation" are all very good indicators of a lead getting converted to an actual paying customer.