

# Lead Case Study Summary

X Education company gets a lot of leads but the leads that convert into paying customers are very few i.e., it is around 30%. The company expects us to build a model and the CEO's expectation is 80% conversion rate.

## **The Following steps were followed to build the model.**

1. Reading and understanding the Data.
2. Data Cleaning
  - The elements of every column were studied carefully and the columns with more null values  $\geq 45\%$  were dropped.
  - The columns with quite a high number of values (nearly 97% of the Data) were dropped.
  - Columns with unique values were dropped.
  - Columns with Nan values were also dropped.
  - Unnecessary columns were dropped.
3. Exploratory Data Analysis.
  - Univariate and Bivariate Analysis for performed.
  - Graphs were plotted of different variables against converted to draw valuable insights.
  - Recommendations and comments were given based on the inferences drawn.
4. Data Preparation
  - Dummy variables were created for categorical variables.
  - Splitting the Data frame into Train and Test sets (70:30).
  - Scaling using Standardization.
  - Dropping a few more columns that were highly correlated to each other.
5. Model Building
  - To make data more manageable, RFE was used to reduce the number of variables.
  - Model building was done thrice, and every time p value was evaluated.
  - After the 3<sup>rd</sup> model building the p values of all the variables was  $< 0.05$ .
  - $VIF < 5$  was also maintained.
6. Model Evaluation
  - Based on the accuracy, specificity and sensitivity graph, 0.38 was chosen as the cut off.
  - The cutoff gave an overall accuracy of 77.78%.

- When we did precision recall view the metrics dropped.
7. Making predictions on test set.
    - Overall accuracy came to 77.098%
    - Sensitivity, specificity, precision, and recall were evaluated.
  8. Conclusions
    - Total time spent on the website is directly proportional to the number of converts and hence the website should be made more engaging and interesting.
    - Total number of visits also has a huge impact on the number of converts.
    - Lead Source, Lead Origin and When was the last activity are also some of the most important variables.
    - The occupation column also provides valuable insights like which kind of people are more likely to be converted to paying customers.