

# **Lead Scoring Case Study**

A brief overview and step by step explanation of the processes followed to conclude potential parameters which might affect conversion of clients.

- I. Data Cleaning -
  - A. Study the data – No of rows and columns, data type, percentage of null values, Outliers presence etc.
  - B. Drop columns which have more than or equal to 35% of null values as it will not contribute much to the analysis and might mislead us if overridden.
  - C. Replacing null values with “Not Provided”. We can deal with it later if required.
  
- II. Exploratory Data Analysis –
  - A. Univariate Analysis of Categorical & Numerical variables.
  - B. Finding relations between Categorical variables and Target variable “Converted”.
  
- III. Data Transformation –
  - A. Creation of dummy variable for Categorical variables.
  - B. Adding dummy variable columns to main data frame.
  - C. Delete categorical columns for which dummy columns have been created.
  
- IV. Data Preparation –
  - A. Splitting the data into test and train dataset.
  - B. Scaling the dataset using MinMaxScaler().
  - C. Plotting a Seaborn Heatmap to check for correlations.

V. Model Building –

- A. Run a Recursive Feature Elimination tool to extract the top 15 parameters that affects the model.
- B. Check Insignificance of the variable (P Value) and VIF to know which variables could further be removed from the analysis.
- C. Repeat the above process of extracting P value and VIF until the P-Value obtained for all parameters is less than 0.05 and VIF is less than 5.
- D. With the help of the final trained model predict values of probability and compare it with actual values.
- E. Define metrics like Accuracy, Sensitivity & Specificity to check strength of the model.
- F. Optimize the cutoff probability value with the help of a ROC curve and again check the values obtained for Accuracy, Sensitivity & Specificity.
- G. Once an optimal value is obtained and difference between Accuracy, Sensitivity & Specificity is within 5% range, move ahead with Test dataset.
- H. Check for Precision and Recall values for both Train and Test data sets.

VI. Conclusion –

- A. Variables that matter the most in finding potential conversions are –
  - 1. Total Visits
  - 2. Total time spent on website
  - 3. Lead Origin – Lead Add Form
  - 4. Lead Sources – Olark Chat, Welingak Website
  - 5. Do Not Email – Yes
  - 6. What is your current occupation? – Other, Student, Unemployed & Working Professional

Keeping this in mind, X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mindset and buy their courses.