Lead Scoring Case Study Summary

Stages in making a model

1. Data Gathering
2. Data Cleaning
3. Exploratory Data Analysis
4. Data Preparation
5. Model Building
6. Model Evaluation
7. Observations

Data Gathering

- Data set – Leads.csv
- Data Dictionary – Leads Data Dictionary.xlsx

Data Cleaning

- In data cleaning stage we verified the percentage of null values and dropped columns having more than 40% of NULL values.
- For the rest of the columns, we worked on each column individually and took necessary to impute its NULL values or dropped the column.

Exploratory Data Analysis

- During Exploratory Data Analysis, we analysed each column individually for both categorical and numerical variables and made some inferences that aligned with business objective.

Data Preparation

- In Data preparation stage, columns having binary values(Yes/No) are converted them to **1/0**
- Also created **dummies variables for categorical** variables
- Also dropped the original categorical variables

Model Building

- In model building stage , we created X and y data sets. **X** set contains all the **independent** variables and **y** set contains the dependent variable i.e. "**Converted**".
- Next step was to split the data set into a train test split (**70**% - train data **30**% – test data)
- Also we scaled the data using **Standard Scaler technique** on necessary variables of our training data set to get their values in scale with all other columns.
- Feature selection was done using Recursive Feature Elimination (**RFE**).

- We verified VIF for the model to confirm that there there is no multicollinearity
- Then **final model was built** based on the features chosen earlier using RFE.
- Based on this model, we also created a prediction and converted probability column and a **confusion matrix** of converted vs predicted.

Model Evaluation

Below is the model Evaluation for the case study

Train Data:

- **Accuracy :** 92.18%
- **Sensitivity :** 91.63%
- **Specificity :** 92.50%

Test Data:

- **Accuracy :** 92.69%
- **Sensitivity :** 90.49%
- **Specificity :** 94.03%

Observations:

It was found that the variables that mattered the most in the potential buyers are (In descending order) :
- The total time spend on the Website.
- Total number of visits.
- When the lead source was:
  - a. Google
  - b. Direct traffic
  - c. Organic search
  - d. Welingak website
- When the last activity was:
  - a. SMS
  - b. Olark chat conversation
- When the lead origin is Lead add format.
- When their current occupation is as a working professional.

Keeping these in consideration the X Education can increase the business