Summary Report

Contributors: Vidhi Surana, Shubham Batch : DS C48

Problem Statement

- Improve the Lead Conversion Rate of the education company called X Education that sells online courses to industry professionals

  Lead Conversion Rate → Leads that convert / Total number of Leads

- When people fill up a form providing their email address or phone number, they are classified to be a lead or a potential customer.
- Now, although X Education gets a lot of leads, its lead conversion rate is very poor. Main problem is how to improve the same.

Business Objective:

- identify the most promising leads(hot leads), i.e. the leads that are most likely to convert into paying customers.
- Assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.
- **The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.**

Solution offered:

Steps followed for solving the problem:

- Data Understanding
- Data Cleaning
  - Handling Select values, replacing them with NULL
  - Handling Missing values and outliers
  - Imputation of Missing values
  - Binary Encoding (1/0)
- Data Preparation
  - Dummy values creation

- Train Test Split
- Perform Scaling – Standard Scaler

➢ Univariate , Bivariate and Multivariate analysis

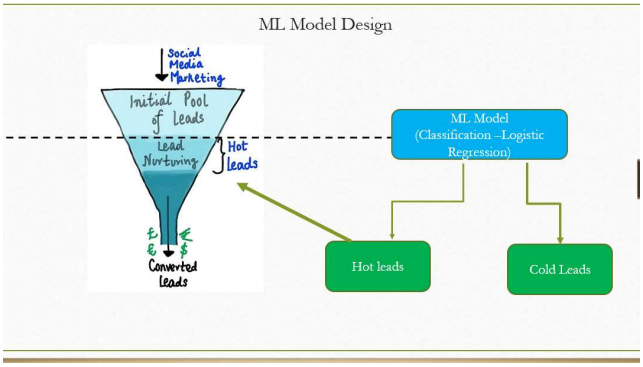
➢ Model Building & Evaluation

- Feature selection using RFE with output number of variables = 20
- Build Logistic Regression Model with good sensitivity
- Evaluating the model with metrics
- Find out the optimal probability Cutoff
- Model diagnosis using ROC Curve, Precision-Recall curve and probability calibration curve
- Use the model prediction on the test dataset and perform model evaluation for the test dataset

➢ Final Step:

- After having LCR > 80%, Assign Lead Score
- Predict Hot leads

Model is positioned in the system as shown below



Analysis:

After bivariate analysis, it was found that for better results,

➢ Focus more on how to increase lead conversion rate from API and Landing Page Submission and also get more leads from Add Form.

➢ Need to focus on the above Lead Sources from Reference and Welingkak Website

- Need to focus on people whose last activity is SMS Sent.
- Need to focus on people who are working professionals.
- Designing the website in such a way that clients spent more time on the website (like can put videos of alumni or quotes) will help.

Model Building:

- After Model Building, the final model had 12 features.
- We selected model with good Sensitivity or Recall to make sure maximum leads are correctly identified as converted
- From the probability Curve, 0.32 came as the optimum point to take it as a cutoff probability.
- The Model seems to have achieved the goal with sensitivity > 80%.
- Selection of the model here is on the basis of Sensitivity/Recall which is above 80%

Making predictions on the test Data set

- Using the probability threshold value of 0.32, the leads from the test dataset were predicted if they will convert or not.
- NOTE : This threshold value can be adjusted to get more sensitivity (decrease the optimal threshold) or get more specificity(increase the optimal threshold)

Lead Score is calculated by the following formula:

- Lead Score for each lead = Lead Conversion Probability for the lead X 100.

Recommendations:

1. The X Education company should be contacting / calling those leads
    - Whose **Lead Source is 'Reference or Wellingak Website**
    - Who are **Working Professionals**.
    - Whose **Last Activity** is **SMS Sent or Others**.
    - Whose **Lead Source** is **'Olark Chat, Google and others'**
    - For whom **Total Time Spent on Website** is high
2. The company should not be calling those leads
    - Who have their **specialization** in Hospital Management.
    - Whose **Last Notable Activity as 'Modified'**.

- Who have chosen the option of **Do Not Email as "yes"**.