Summary

X Education, an online course provider in the edtech industry, faced a challenge of low lead conversion rates despite generating a substantial number of leads (30%). The company aimed to enhance its lead conversion process by identifying potential high-converting leads, referred to as "Hot Leads," with the ultimate goal of increasing the overall conversion rate to 80%. To achieve this, the company embarked on a comprehensive case study involving data analysis, cleaning, exploratory data analysis (EDA), data preparation, model building, evaluation, and providing actionable recommendations.

Data Inspection and Data Cleaning

In the initial stages of the case study, a thorough data inspection was conducted, encompassing preliminary checks, manipulation of data types, distribution analysis, and checks for missing values and outliers. The subsequent data cleaning process involved dropping columns with single values, those related to customer IDs, and columns with over 35% null values. Categorical variables with 'select' categories were converted to null values, and certain categories contributing minimally were grouped for more meaningful analysis.

Exploratory Data Analysis

There is a data Imbalance in the data given as there is only 37.5% conversion rate. We did univariate and Bivariate Analysis. This included count plots for categorical variable and histogram for numerical variables. In bivariate analysis we tried to analyze each column with respect to target using count plots. To look in correlations between different variables we took help from correlation heatmap. EDA also helped us to find out which numerical column had outliers.

Data Preparation

Created dummy values for categorical features then split the data into Train & Test Sets in 70:30 ratio. We also did feature Scaling using MinMaxScaler.

Model Building

We used Recursive Feature Elimination to reduce variables to 15. This made Dataframe more manageable. Manual Feature Reduction process was used to build models by dropping variables with p value > 0.05. Total 3 models were built before reaching final Model 4 which was stable with (p-values < 0.05). There were no signs of multicollinearity with VIF < 5.

Model Evaluation(On train and test data)

Confusion matrix was made and cut off point of 0.35 was selected based on accuracy, sensitivity and specificity plot. This cut off gave accuracy, specificity and sensitivity all near to 80%. Whereas precision recall view gave less performance metrics around 72%. So as to solve business problem CEO asked to boost conversion rate to 80%, we gave weightage to sensitivity, specificity view, so our optimal cut-off for final predictions for Lead score was assigned to train data using 0.35. A similar kind of model performance was observed on test data as well when a cutoff of 0.35 was implemented on the data.

Recommendations

The recommendations derived from the regression model highlighted key variables influencing lead conversion. Notably, factors such as total time spent on the website, total visits, lead source from the company's website, last notable activity (e.g., email opened, chat conversation, page visits), lead origin (specifically, lead add form), current occupation (especially working professionals), and pages viewed per visit emerged as significant contributors to the likelihood of conversion. In conclusion, X Education can maximize profits by strategically focusing on leads exhibiting these influential characteristics. By incorporating these findings into their sales and marketing strategies, the company can enhance lead targeting, improve overall efficiency, and achieve the desired conversion rate of 80%.