Manasa Manjunatha
UpGrad - IIITB
EPGD Data Science
Batch Jan 2024

Lead Scoring Case Study Summary

23rd July 2024

Problem Statement:

X Education, which sells online courses to industry professionals, needed a model to identify and prioritize leads most likely to convert into paying customers. The goal was to achieve a lead conversion rate of approximately 80%.

Solution Summary:

1. Data Understanding and Cleaning:

- Analyzed and cleaned the data, dropping variables with high NULL values and imputing missing values. Created new classification variables for categorical data and removed outliers.

2. Exploratory Data Analysis:

- Conducted an analysis to understand data orientation, dropping variables with only one unique value.

3. Feature Engineering:

- Created dummy variables for categorical data and applied Min-Max Scaling to numerical variables.

4. Data Splitting:

- Divided the data into training (80%) and test (20%) sets.

5. Model Building and Feature Selection:

- Used Recursive Feature Elimination to select the top features, refining to 21 based on statistical significance. Built a model to predict conversion probabilities with a cutoff of 0.5.

6. Model Evaluation with Test Set Predictions:

- Evaluated the model in terms of accuracy, sensitivity, precision and specificity metrics, by applying the model to the test set, achieving 82% accuracy, 69% sensitivity, 81% precision and 90% specificity.