Lead Scoring Case Study

Summary

This study is conducted on behalf of X Education with the objective of attracting more industry professionals to enroll in their courses. The essential data provided has offered valuable insights into the browsing behavior of potential customers, encompassing site visit patterns, engagement duration, referral sources, and the overall conversion rate.

The steps taken are outlined as follows:

1. Read and Comprehend the Data:

- Load the dataset, analyze its structure, and understand its contents.

2. Data Cleaning:

- Identify and handle null values and specific values, replacing designated values with "nan" to signify missing information.
- Remove columns with a high percentage of null values.
- Impute missing values using median/mode for numerical variables.
- Create new categorical variables for categorical features as needed.

3. Exploratory Data Analysis (EDA):

- Conduct a detailed analysis, revealing irrelevant elements within categorical variables.

- Numeric values appeared satisfactory with no outliers detected.

4. Dummy Variables:

- Create dummy variables to represent categorical values in the dataset.

5. Train-Test Split:

- Divide the dataset into training and testing sets, with a 70% distribution for training and 30% for testing.

6. Feature Scaling:

- Apply MINMAXScaler to normalize the original numerical values in the dataset.

7. Model Building:

- Implement Recursive Feature Elimination to identify the top 15 most relevant variables.

- Remove remaining variables based on VIF values and p-values.

8. Model Evaluation:

- Generate a confusion matrix.

- Determine the optimal cutoff value using the ROC curve to calculate accuracy, sensitivity, and specificity, achieving approximately 80% accuracy.

9. Prediction:

- Utilize an optimal cutoff of 0.35 for predictions on the test dataframe, achieving 80% accuracy, sensitivity, and specificity.

Discussion:

To enhance lead conversion during the intern-hiring period, X Education can adopt a targeted strategy by focusing on high-potential leads from the Welingak Website and reference sources. Prioritizing calls to these leads and utilizing effective communication channels, such as SMS and email, can improve conversion rates. Additionally, prioritizing leads based on website engagement can further increase the likelihood of success.

To minimize unnecessary phone calls post-achieving quarterly sales targets ahead of schedule, the sales team can shift focus to lead nurturing activities like personalized emails, SMS messages, and targeted newsletters. Automation of SMS messages for high-conversion probability customers can optimize outreach efforts. Seeking feedback from existing customers helps refine lead quality and conversion rates. Collaboration among sales, management, and data scientists is crucial to continually improve the lead

conversion model. Developing a strategy for discounts or incentives can motivate potential customers to take prompt action. This comprehensive approach ensures efficient lead conversion and minimizes unnecessary outreach efforts after reaching sales targets.