

Lead Scoring Case Study

Summary:

This research was carried out on behalf of X Education with the objective of enticing a greater number of industry professionals to register for their courses. The essential information gathered has provided valuable observations into the online browsing habits of prospective customers, encompassing patterns of site visits, duration of engagement, referral sources, and the overall rate of conversion.

The following steps were employed in the process:

1. **Data Reading and Comprehension**: The dataset was loaded, its contents were examined, and its structure or shape was analyzed.
2. **Data Cleansing**: Detection and management of null values and specific entries were carried out. Designated values were replaced with "nan" to indicate missing information. Columns with a high proportion of null values were eliminated, and missing values in numerical variables were imputed using median/mode. Additionally, new categorical variables were generated for necessary categorical features.
3. **Exploratory Data Analysis (EDA)**: A comprehensive analysis was conducted, revealing that several elements within categorical variables were irrelevant. Numeric values were found to be satisfactory, showing no outliers.
4. **Creation of Dummy Variables**: Dummy variables were established to represent categorical values within the dataset.
5. **Train-Test Split**: The dataset was partitioned into training and testing sets, with a distribution ratio of 70% for training and 30% for testing.
6. **Feature Scaling**: The MINMAXScaler was applied to normalize the original numerical values present in the dataset.
7. **Model Building**: Utilized Recursive Feature Elimination to pinpoint the most significant 15 variables. Following this, eliminated the remaining variables based on their VIF values and p-values.

8. Model Evaluation: Generated a confusion matrix and then employed the optimal cutoff value, determined through the ROC curve, to calculate accuracy, sensitivity, and specificity. The outcome was an accuracy of around 80%.

9. Predictions: Employed an optimal cutoff of 0.35 for predictions on the test data set, resulting in an accuracy, sensitivity, and specificity of 80%.

Discussion:

To improve lead conversion during the intern hiring phase, X Education can implement a targeted approach by concentrating on high-potential leads, particularly those sourced from the Welingak Website and referrals. The sales team should give priority to contacting these leads and utilize effective communication channels, such as SMS and email, to enhance conversion rates. Additionally, prioritizing leads based on their engagement with the website can further boost the chances of success.

To reduce unnecessary phone calls after surpassing quarterly sales targets ahead of schedule, the sales team can shift their focus to lead nurturing activities, such as personalized emails, SMS messages, and targeted newsletters. Automating SMS messages for customers with a high likelihood of conversion can optimize outreach efforts. Soliciting feedback from existing customers is crucial for refining lead quality and conversion rates. Collaboration among the sales team, management, and data scientists is essential to continuously enhance the lead conversion model. Developing a strategy for discounts or incentives can encourage potential customers to take immediate action. This comprehensive approach ensures effective lead conversion and minimizes unnecessary outreach efforts once sales targets are achieved.