

Summary Report: Enhancing Lead Conversion Rates with a Scoring Model

Problem Statement

X Education aims to improve lead conversion rates in its online courses business by developing a lead scoring model. The goal is to identify promising leads with a higher likelihood of conversion. This involves utilizing historical data, model training, setting conversion thresholds, and ongoing refinement.

Steps Followed

1. **Importing Libraries and Data**
2. **Data Understanding and Inspection**
3. **Data Cleaning:** Removed columns with $\geq 40\%$ missing values, treated skewness, imputed missing values, and handled outliers.
4. **EDA:**
 - Univariate analysis revealed insights about various features' distributions.
 - Bivariate analysis highlighted correlations between features and lead conversion.
5. **Data Preparation:** Created dummy variables for categorical features.
6. **Test-Train Split**
7. **Feature Scaling:** Applied MinMaxScaler to numeric variables.
8. **Feature Selection:** Used recursive feature elimination (RFE) to select 15 important variables.
9. **Model Building:** Developed and evaluated five models based on p-value and VIF criteria.
10. **Model Evaluation:** Optimal cutoff determined for predictions based on sensitivity-specificity view.
11. **Prediction on the Test Set:** Performance metrics like accuracy, precision, recall were computed.
12. **Conclusion:**
 - Train Set:
 - Accuracy: 78.79%
 - Sensitivity: 77.9%
 - Specificity: 79.31%
 - Test Set:
 - Accuracy: 78.72%
 - Sensitivity: 76.89%
 - Specificity: 79.99%

- Model performed consistently across different metrics in both train and test datasets.
- Model's sensitivity is close to the target of 80% set by the CEO.

Top 3 Features Contributing to Predicting Hot Leads:

1. **Lead Source_Welingak Website:** This source significantly impacts lead conversion.
2. **Total Time Spent on Website:** Longer engagement positively correlates with conversion.
3. **Occupation_Working Professional:** Leads with this occupation show a higher likelihood of conversion.

Business Recommendations:

To Increase Lead Conversion Rates:

1. Prioritize features with positive coefficients to enhance targeted marketing precision.
2. Focus on top-performing lead sources to attract high-quality leads.
3. Tailor communication to connect effectively with employed professionals.
4. Optimize communication channels based on lead engagement impact.
5. Allocate budget for activities on the Welingak Website to leverage its influence.
6. Introduce incentives for successful references to boost lead submission.
7. Target working professionals due to their high conversion rates and better financial capacity.

To Identify Areas of Improvement:

1. Evaluate specialization offerings with negative coefficients.
2. Review the landing page submission process for potential enhancements.

The developed logistic regression model with 11 features proved effective in identifying promising leads for conversion. By focusing on the recommended strategies and refining the model based on ongoing data, X Education can expect to achieve its goal of improving lead conversion rates.