

Summary Report: Logistic Regression Model for Lead Scoring

As part of our engagement with X Education, we were tasked with developing a logistic regression model to assign a lead score (0-100) that accurately identifies "Hot Leads" likely to convert into paying customers.

Business Objective

Our objective was to improve the company's lead conversion rate by identifying high-potential leads and optimizing the sales process.

Methodology

1. Data Cleaning

We performed thorough data cleaning, dropping columns with >70% nulls, imputing numerical categorical data with mode, and treating outliers.

2. Exploratory Data Analysis (EDA):

We conducted univariate and bivariate analysis to identify key variables influencing the target variable. Notably, we found data imbalance in the target variable and performed further analysis on categorical and numerical variables.

3. Data Preparation

We created dummy features for categorical variables, split our dataset into train and test sets (70:30 ratio), and applied feature scaling using standardization.

4. Model Building

We built a vanilla logistic regression model and a second model using hyperparameters and GridSearchCV to optimize the model's performance.

5. Model Evaluation

We evaluated both models using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The second model performed slightly better over the test data, and we selected it as the final model.

Key Findings

Our analysis revealed that a combination of demographic, behavioural, and lead-related variables are strong predictors of lead conversion as mentioned below,

1. The total Time Spent on the Website is the most influential feature in predicting lead conversion.
2. Occupation - Working Professional has a significant positive impact, indicating higher conversion rates among working professionals.
3. Last Activity - SMS Sent and Email Opened are important indicators of lead conversion.
4. Missing Occupation Information and Specialization Information surprisingly contribute positively to lead conversion.

We believe our model can help X Education increase its lead conversion rate by targeting high-potential leads and optimizing the sales process.

Recommendations

1. **Focus on High-Engagement Channels:** Enhance and optimize channels like SMS, email, and Olark Chat (live chat) for higher conversion rates.
2. **Leverage Time Spent on Website:** Develop strategies to increase user engagement time on the website through engaging content and interactive elements.
3. **Occupation-Based Targeting:** Target working professionals more aggressively since they are more likely to convert.
4. **Form Origin Optimization:** Prioritize leads generated from specific forms like Lead Add Form and Landing Page Submissions.
5. **Handle Missing Data Strategically:** Consider that missing information in certain fields (Occupation and Specialization) might still lead to high conversion rates, indicating other underlying factors at play.
6. **Utilize Geographical Insights:** Customize marketing strategies for leads from regions like Saudi Arabia and Bahrain to maximize conversion rates.

By following this approach, we believe X Education can achieve its goal of improving its lead conversion rate and driving business growth.