Summary Report on Lead Conversion Optimization for X Education

Project Overview

X Education, an online course provider, aimed to improve its lead conversion rate from 30% to 80% by identifying 'Hot Leads.' The project involved developing a logistic regression model to assign lead scores based on key attributes, enabling targeted communication and more efficient resource allocation for sales efforts.

Data Processing and Preparation

The initial dataset comprised over 9,240 rows and 37 columns. The first step involved checking for duplicates, which confirmed the uniqueness of all entries. The focus then shifted to handling missing values. Columns with a high percentage of missing values, particularly those not deemed essential for the analysis, were removed. Specifically, columns such as 'City' and 'Country' were dropped due to their lack of relevance to an online platform with a global audience. Other columns with predominantly single values, like 'Do Not Call' and various marketing channels, were also excluded to avoid redundancy in insights. The data was then cleaned, ensuring that only relevant and informative attributes remained.

Model Development

The core objective was to create a logistic regression model that assigns a lead score ranging from 0 to 100. This score would indicate the likelihood of conversion, allowing the sales team to prioritize high-potential leads. Various attributes such as 'Total Time Spent on Website', 'Lead Source', and 'Last Activity' were included as predictors. The model was trained on the cleaned dataset, and statistical significance was evaluated for each variable.

Model Evaluation and Performance

Two evaluation metrics were employed to assess the model's performance: the AUC ROC curve and the Precision-Recall curve. The model achieved an accuracy of 79% with the ROC AUC cut-off and slightly improved to 80% when evaluated with the Precision-Recall curve. Precision and recall metrics highlighted a balanced trade-off between false positives and false negatives, enabling adaptability to different business scenarios. The findings suggested that while a focus on precision minimized unnecessary marketing efforts, it did come at the cost of potentially overlooking some conversion opportunities.

Key Learnings

1. **Data Quality Matters:** Thorough data cleaning and preparation are crucial in enhancing model accuracy. Columns with excessive missing values or low variance can lead to misinterpretation of the data and should be scrutinized before analysis.
2. **Variable Significance:** The selection of variables plays a pivotal role in model performance. Identifying key predictors that significantly influence conversion can streamline sales efforts and improve efficiency.
3. **Flexible Model Deployment:** The ability to adjust the model’s evaluation criteria based on changing business objectives (e.g., focusing on precision vs. recall) provides a valuable tool for X Education to navigate varying marketing strategies.
4. **Continuous Improvement:** The project's iterative nature highlights the importance of refining the model based on real-world outcomes and feedback, ensuring that it evolves alongside business needs.

Conclusion

This project successfully developed a robust logistic regression model that can help X Education enhance its lead conversion strategy. By focusing on high-potential leads, the company can allocate resources more effectively and work towards achieving its goal of an 80% conversion rate. Future efforts should emphasize continuous data monitoring and model refinement to maintain alignment with dynamic market conditions.