Summary Report: Lead Scoring for X Education

Ann Reji Thomas//Jibin Baby//Poornima

May 26, 2024

1 Introduction

For the assignment focused on lead scoring for X Education, this report provides a brief summary of the approach taken, the methodologies employed, and the key learnings gathered throughout the process.

2 Understanding the Problem Statement

The problem statement outlined the challenges faced by X Education in converting leads into paying customers. The goal was to build a model that could predict the likelihood of conversion for each lead, thereby enabling the sales team to prioritize high-potential leads.

3 Data Exploration and Preprocessing

I began by exploring the provided dataset, which contained various attributes such as lead source, total time spent on the website, and more. I identified missing values and handled them appropriately. Additionally, I encoded categorical variables and scaled numerical features to prepare the data for modeling.

4 Model Selection and Training

Given the binary nature of the target variable (converted or not converted), I chose logistic regression as the algorithm for modeling. Logistic regression is a widely used technique for binary classification tasks and is known for its simplicity and interpretability. I trained the logistic regression model using the preprocessed data.

5 Model Evaluation

To evaluate the performance of the model, I utilized various metrics such as accuracy, precision, recall, and the ROC-AUC score. These metrics provided insights into how well the model performed in distinguishing between converted and non-converted leads. Additionally, I visualized the confusion matrix to gain a deeper understanding of the model's predictions.

6 Lead Scoring and Implementation

One of the key aspects of the assignment was the implementation of lead scoring. Using the trained logistic regression model, I assigned lead scores to each lead in the dataset based on their likelihood of conversion. This facilitated the identification of high-potential leads, allowing the sales team to focus their efforts more effectively.

7 Key Learnings

Throughout the assignment, I gained several valuable learnings:

1. **Data Preprocessing Techniques:** I learned various techniques for handling missing data, encoding categorical variables, and scaling numerical features, which are essential steps in preparing data for modeling.

2. **Model Evaluation Metrics:** I gained a deeper understanding of metrics such as accuracy, precision, recall, and ROC-AUC score and their significance in evaluating the performance of classification models.

3. **Practical Implementation:** The assignment provided hands-on experience in building and implementing a predictive model for real-world business problems, thereby enhancing my practical skills in data science.

4. **Importance of Lead Scoring:** I learned about the importance of lead scoring in optimizing sales and marketing efforts, and how predictive modeling can be leveraged to prioritize leads effectively.

8 Conclusion

The assignment was a valuable learning experience that helped me strengthen my skills in data preprocessing, model building, and evaluation. It provided insights into the practical application of machine learning techniques in solving business problems and highlighted the importance of data-driven decision-making processes.