Lead Scoring Case Study - Comprehensive Summary Report

Overview

This report presents an in-depth analysis of the lead scoring case study, focusing on data preprocessing, exploratory data analysis (EDA), model building, and evaluation. It integrates strategies for lead conversion optimization and insights derived from the lead scoring model.

Key Insights and Findings

1. Data Overview and Understanding:

- Initial data: 9240 records with 37 columns (30 categorical, 7 numerical).

- Objective: Improve lead conversion through effective scoring and modeling.

2. Data Cleaning and Preparation:

- Replaced invalid class values (e.g., 'Select') with NaN.

- Dropped columns with only one unique value or >40% missing values.

- Bucketed high-class categorical variables for better representation.

- Treated missing values with domain knowledge (e.g., added 'Not Disclosed' for Specialization).

- Removed outliers (2.8% of data) and split dataset into Train (70%) and Test (30%).

3. EDA and Visualizations:

- Visualizations included box plots, pair plots, and count plots for key variables.

- Correlation heatmap used to identify redundant features.

- Insights from visualizations integrated into modeling.

4. Feature Engineering and Model Building:

- Used RFE to select top 16 features.

- Built and refined seven Logistic Regression models, iteratively eliminating features based on p-values and VIF values.

5. Model Insights and Metrics:

- Training Data (cutoff 0.5):

  - Accuracy: 80.91%

  - Sensitivity: 79.94%

  - Specificity: 81.50%

- Optimal probability cutoff found at 0.32.

- Lead scoring implemented: Probability × 100 to rank leads.

Strategies for Specific Scenarios

1. Aggressive Conversion During Internship Phase:

- Segment high-priority leads (score >80).

- Employ real-time tracking and proactive follow-ups.

- Personalize communication to optimize engagement.

2. Minimizing Efforts Post-Target Achievement:

- Focus only on leads with scores >90.

- Automate low-priority communication using email/SMS campaigns.

- Adjust thresholds dynamically based on quarterly performance.

Visual Insights

This report is supplemented with the following visualizations:

1. Top Variables Bar Chart: Importance of predictive features.

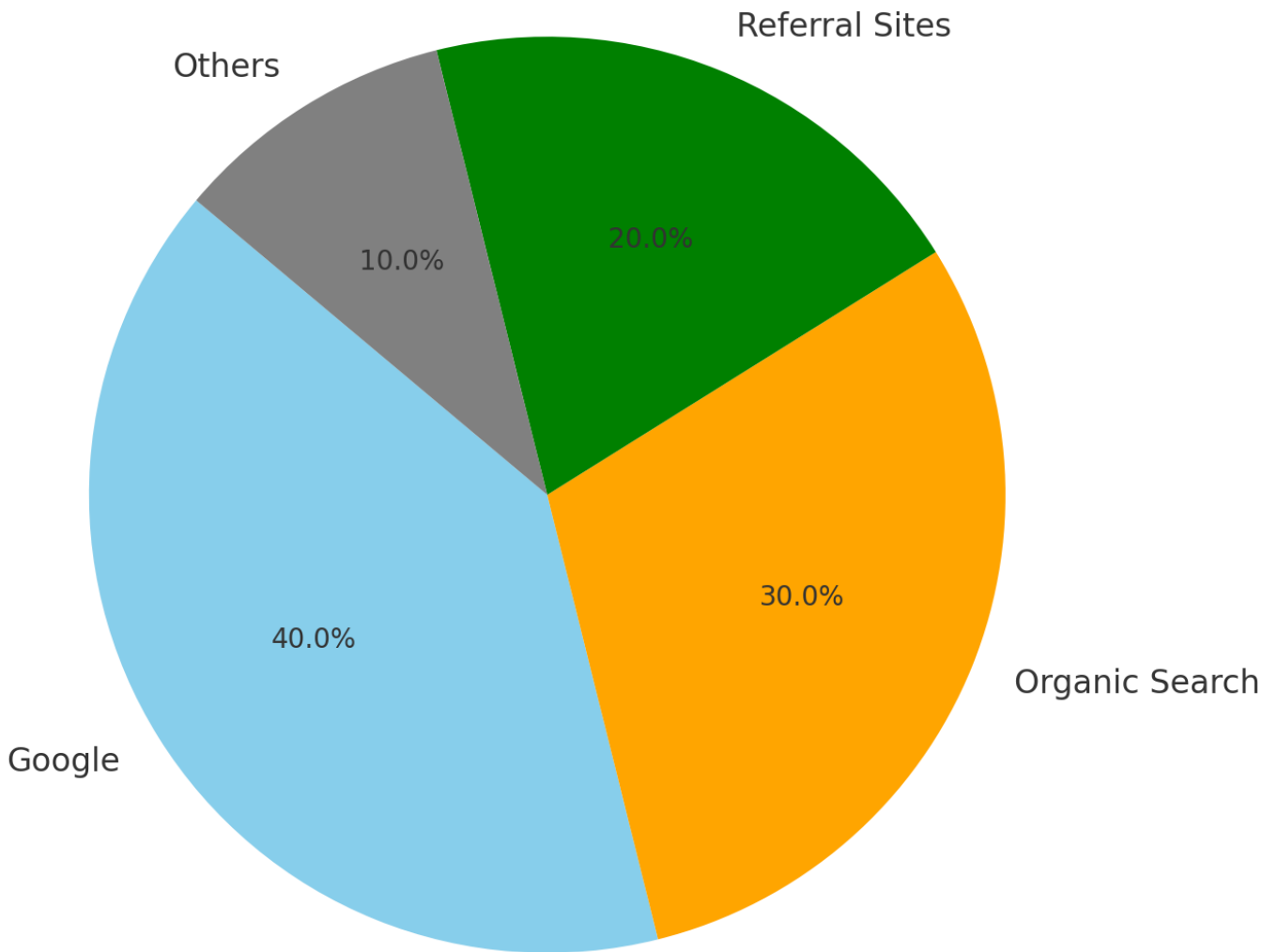2. Lead Source Distribution Pie Chart: Proportions of leads by source.

3. Evaluation Metrics Line Plot: Accuracy, precision, and recall across cutoff values.

Top Variables Contributing to Lead Conversion

Top Variables Contributing to Lead Conversion
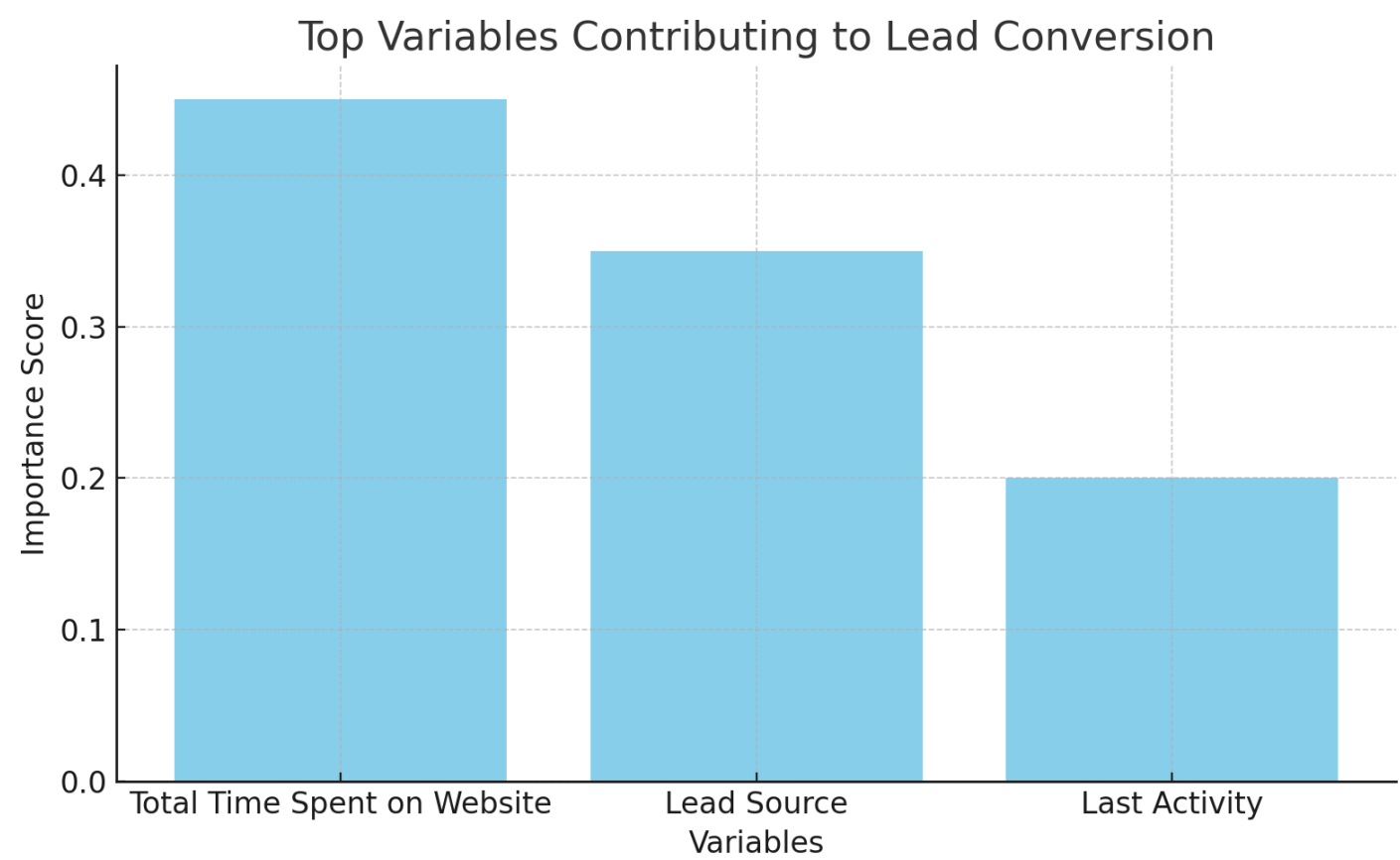
Lead Source Distribution

Lead Source Distribution

Model Evaluation Metrics at Different Cutoff Values

Model Evaluation Metrics at Different Cutoff Values