Enhancing Efficiency in Lead Management

Lead Conversion Optimisation Analysis for X Education

Introduction

Problem Statement:

- X Education struggles with a low lead conversion rate of 30%, despite high potential interest on their platform.

Objective:

- The goal is to enhance lead conversion efficiency by developing a predictive model to identify and prioritise 'Hot Leads', aiming to achieve an 80% conversion target.

Data Overview

Dataset Description:

- The dataset contains approximately 9,000 data points, with features such as Lead Source, Total Time Spent on Website, Total Visits, and Last Activity.

- The target variable is 'Converted', indicating whether a lead was successfully converted (1) or not (0).

Initial Observations:

- The dataset includes a mix of numerical and categorical variables.

- Early data cleaning steps included removing identifiers such as Prospect ID and Lead Number, which are not useful for predictive modelling.

Data Cleaning and Preparation

Data Cleaning Steps:

- **Missing Values**: Addressed by imputing or removing, depending on the significance and the amount of missing data.

- **Outliers**: Managed by identifying and capping extreme values based on quantile thresholds.

Feature Engineering:

- **New Features:** Enhanced model predictability by deriving new variables from existing data.

- **Categorical Variables:** Transformed into dummy variables for regression, avoiding multicollinearity by excluding the first category in each set.

Exploratory Data Analysis (EDA)

Key Visualisations:

- **Distribution of 'Converted' Variable**: Visual representation highlighting the proportion of converted versus non-converted leads.

- **Conversion Rates by Lead Origin and Source**: Illustrated the significant impact these features have on lead conversion.

Insights Gained:

- **High-Value Origins and Sources**: Identified specific origins and sources with higher conversion rates, suggesting targeted marketing strategies.

- **Engagement and Conversion**: Time spent on the website and page visits correlated with higher conversion chances, underscoring their importance in lead qualification.

Model Building

Model Selection:

- **Logistic Regression**: Chosen for its ability to provide probabilities for outcome prediction, facilitating the assignment of lead scores based on the likelihood of conversion.

Model Training:

- **Data Split**: Utilised 80/20 train-test split to ensure model robustness and avoid overfitting.

- **Training Process**: Model trained on cleaned and prepared data, focusing on features identified as significant during EDA.

Model Evaluation

Performance Metrics:

- **Accuracy**: Evaluated the overall accuracy of the model to predict conversion.

- **Precision and Recall**: Focused on the precision to minimize false positives and recall to capture as many true positives as possible.

- **ROC-AUC Score**: Utilized ROC curve and AUC score to measure the model's ability to discriminate between converted and non-converted leads.

Results Summary:

- Display key metrics like accuracy, precision, recall, and the AUC score.

- Discuss the balance between precision and recall, highlighting the trade-offs and the chosen optimization based on business priorities.

Lead Scoring System

Conversion of Model Output to Lead Scores:

- **Probability to Score Conversion**: Transformed model probabilities into scores from 0 to 100, enabling sales teams to prioritize leads based on their likelihood of conversion.

- **Dynamic Scoring**: Scores are updated dynamically to reflect new interactions, ensuring continuous relevance and accuracy.

Impact on Sales Efficiency:

- **Prioritization and Resource Allocation**: High-scoring leads are prioritized, optimizing the sales team's focus and efforts, which has significantly improved conversion rates.

- **Statistical Improvement**: Highlighted key performance improvements, such as increased conversion rates and more efficient lead management, demonstrating the scoring system's effectiveness.

Strategic Recommendations and Business Impact

Strategic Recommendations:

- **Optimize High-Impact Channels**: Enhance focus on lead sources with the highest conversion rates, such as 'Lead Add Form' and 'Welingak Website'.

- **Tailored Communication Plans**: Utilize lead scores to implement targeted engagement strategies for high-potential leads to boost conversions.

Business Impact:

- **Efficient Lead Management**: The lead scoring system improves prioritization and resource allocation, leading to more effective lead nurturing and higher conversion rates.

- **Integration and Improvement**: Recommend ongoing model refinements and integration with marketing tools to sustain and enhance lead management efficiencies.