CASE STUDY- LEAD SCORING

Executive Summary

This analysis was conducted for X Education with the objective of optimizing strategies to attract more industry professionals to enrol in their courses. The initial dataset provided valuable insights into the behaviour of potential customers, encompassing their site visits, duration spent, referral sources, and the overall conversion rate.
The analytical approach involved several key steps:

Data Cleaning:

Initial data cleaning addressed converting Boolean values to integer 0/1, null values, and the 'select' option was replaced with null for enhanced interpretability. Null values were standardized to 'not provided' to minimize data loss. Categorical variables were further refined, consolidating elements for improved clarity.

Exploratory Data Analysis (EDA):

A brief EDA was performed to assess the data's condition. Irrelevant elements in categorical variables were identified, while numeric values exhibited sound characteristics with no discernible outliers.

Dummy Variables:

Dummy variables were created, with subsequent removal of dummies containing 'not provided' elements. For numeric values, the MinMaxScaler was applied.

Train-Test Split:

The dataset was partitioned into training (70%) and testing (30%) sets.

Model Building:

Recursive Feature Elimination (RFE) was employed to select the top 15 relevant variables. Subsequently, manual removal of non-essential variables based on VIF values and p-values was conducted (retaining variables with VIF < 5 and p-value < 0.05).

Model Evaluation:

A confusion matrix was generated, and an optimal cut-off value (determined through ROC curve analysis) yielded an accuracy, sensitivity, and specificity of approximately 80%.

Prediction:

Predictions were made on the test dataset using an optimal cut-off of 0.35, resulting in an accuracy, sensitivity, and specificity of 80%.

Precision-Recall Analysis:

A precision-recall approach was employed, revealing a cut-off of 0.41 with precision around 73% and recall around 76% on the test dataset.

Key Insights:

The variables identified as most influential in predicting potential buyers, ranked in descending order of importance, are:

- Total time spent on the website.
- Total number of visits.
- Lead source, with emphasis on Google, Direct traffic, Organic search, and Welingak website.
- Last activity, particularly SMS and Olark chat conversation.
- Lead origin as Lead add format.
- Current occupation as a working professional.

In conclusion, focusing efforts on these key variables presents an opportunity for X Education to significantly enhance its conversion rates by effectively engaging and persuading potential buyers.