SUMMARY REPORT

LEADS SCORING CASE STUDY

X Education, an education company, aims to improve its lead conversion rate by identifying leads with higher potential for active results. In this case study, we analyze the provided leads data, which includes various factors such as total time spent, lead origins, sources, requests to call or email back, specializations, occupations, country, and marketing remarks.

We begin by cleaning the data, addressing null values, and dropping or replacing columns as needed. Our exploratory data analysis reveals that specializations in finance management, banking, investment, and insurance show promise, and occupations such as unemployed and working professionals are more likely to opt for career-focused courses. Flexibility and convenience are also important factors for customers in choosing a course.

Next, we build a machine learning logistic regression model using a 70-30 train-test split, standardizing continuous variables, and selecting features using Recursive Feature Elimination (RFE) with a set of 25 variables. We then remove features with p-values higher than 0.5 and assess multicollinearity using Variance Inflation Factor (VIF). After eliminating features with VIF higher than 3, we arrive at a final model.

We evaluate the model using accuracy, sensitivity, specificity, precision, and recall metrics, with a cutoff of 0.3. The model achieves an accuracy of 90%, sensitivity of 96%, specificity of around 88%, precision score of 83%, and recall score of 96%, indicating good predictive performance.

Based on our analysis, we recommend that the company focuses on leads that show intent to revert back with an email, leads closed by 'Horizzon', and avoids leads that are marked as busy, switched off, not interested in further education, or are already students.

In conclusion, this case study highlights the use of logistic regression for predicting lead conversion and provides insights and recommendations for X Education to optimize their lead conversion rate. The full analysis, including code and results, is available in this repository for reference and further analysis. The model developed through our analysis has demonstrated a high accuracy rate of 91% in predicting lead outcomes. We believe that this level of accuracy will enable the company to make more informed decisions when placing future calls and result in a higher number of converted leads.