Summary

Overview: -

The purpose of this case study was to create a lead scoring model for X Education, an educational online course provider, to enhance their lead conversion rate from 30% to the desired 80%. The aim was to single out high-potential leads so that the sales department would target only the most likely customers instead of calling all the leads randomly.

Process Followed: -

The below systematic process was employed to construct the model:

1. Data Understanding & Cleaning

- Imported and analysed the dataset (9000+ data points).
- Dealt with missing values and dropped irrelevant data.
- Converted categorical variables with dummy encoding.

2. Exploratory Data Analysis (EDA)

- Univariate Analysis: Detected major trends in variables like Lead Source, Total Time Spent on Website, and Lead Profile.
- Bivariate Analysis: Detected significant correlations between Total Time Spent on Website & Conversion Rate and Page Views per Visit & Total Visits.

3. Feature Engineering & Selection

- Chosen the most significant numerical and categorical variables.
- Constructed dummy variables for categorical variables.

4. Model Building & Evaluation

- Applied Logistic Regression for binary classification.
- Assessed the model with Accuracy, Precision, Recall, Sensitivity, and Specificity.
- Tunned the model's cutoff probability to 0.37, finding a balance between precision and recall.
- Final Model Performance:

- Accuracy: 91% (Test Data)

- Sensitivity: 88%

- Specificity: 94%

- These findings suggest that the model can consistently distinguish between potential conversions and non-conversions.

Key Learnings & Insights: -

1. Top Three Numerical Variables Impacting Lead Conversion

- Total Time on Website: A good measure of interest; greater time corresponds to a greater likelihood of conversion.

- Number of Visits: Leads who visit more than once indicate high intent.

- Lead Origin: Where a lead comes from (e.g., referrals versus ads) impacts conversion likelihood quite a bit.

2. Top Three Categorical Variables to Pay Attention To

- Lead Source: Google and LinkedIn produce high-converting leads.

- Last Activity: Leads who open mail or get SMS notifications have a higher chance of conversion.

- Lead Quality: Internal scoring, if available, can enhance the process of conversion.

Business Strategies Based on Model Findings: -

Aggressive Lead Conversion Strategy (Intern Hiring Period)

- Decrease the lead score cutoff (e.g., 0.5 rather than 0.7) to catch more prospective leads.

- Call more frequently and follow up to interact with all high-potential leads.

- Utilize multi-channel outreach (calls, SMS, email) for optimal interaction.

Reducing Calls When Targets Are Achieved

- Raise the lead score cutoff (e.g., 0.8 rather than 0.5) to eliminate weaker leads.

- Utilize automation (emails & SMS) to pre-screen leads prior to calling.

- Prioritize high-revenue leads to guarantee profitability.

Conclusion: -

This case study gave excellent insights into lead behaviour, conversion factors, and sales efficiency strategies. The logistic regression model was able to accurately predict high-converting leads, with a 91% accuracy and good specificity and sensitivity. With the proposed strategies, X Education can dramatically increase its lead conversion rate and sales efficiency.