

## Summary Report: Lead Conversion Prediction Model

The goal of this assignment was to develop a predictive model to determine factors that most influence lead conversion for X Education, aiming to improve conversion rates and efficiently allocate sales resources. The objective was to:

1. Identify the leads most likely to convert.
2. Optimize sales efforts based on conversion probabilities.
3. Develop strategies for high and low conversion phases, especially during intern-heavy periods and post-target achievement times.

### Problem Statement

X Education needed a solution to predict which leads are most likely to convert, allowing the sales team to focus on those with the highest conversion probabilities. The company also wanted to know how to adjust its approach during specific periods, such as when interns are onboarded or when quarterly targets have been met ahead of schedule.

### Approach

#### 1. Data Preprocessing:

- Missing values were handled by either imputation or removal, ensuring the data was complete for analysis.
- Categorical variables were encoded using dummy variables, which allowed for the inclusion of qualitative features in the logistic regression model.
- Numerical variables were scaled to standardize the input features.
- The final dataset contained a mix of categorical and continuous variables, all preprocessed and ready for model training.

#### 2. Model Development:

- A **logistic regression** model was built to classify leads into two categories: those likely to convert (1) and those unlikely to convert (0).
- The dataset was split into training (70%) and test (30%) sets. The model was trained on the training data and evaluated on the test set.
- Evaluation metrics included **accuracy, precision, recall, F1-score**, and the **ROC-AUC score**.
- The model performed well, achieving the following results:
  - **Accuracy:** 94%
  - **Precision:** 92%
  - **Recall:** 95%
  - **F1-Score:** 0.93
  - **ROC-AUC score:** 0.98, indicating a strong ability to distinguish between lead conversions and non-conversions.

#### 3. Model Evaluation:

- The logistic regression model identified the top contributing factors to lead conversion probabilities:
  - **Tag\_Will revert after reading the email** (coefficient = 1.45): This tag increased the likelihood of conversion by 45%.
  - **Lead Profile\_Potential Lead** (coefficient = 1.31): A positive indicator, showing leads in this category were more likely to convert.

- **Lead Quality\_Worst** (coefficient = -1.25): Leads marked with poor quality were less likely to convert, reducing the probability by 25%.

## Key Insights and Learnings

### 1. Top Variables Influencing Conversion:

- The most important variables for lead conversion were identified as lead tags, lead profiles, and lead quality. Specifically, the tag "Will revert after reading the email" and profile type "Potential Lead" were strong positive predictors, while low-quality leads had a much lower conversion probability.

### 2. Intern Phase Strategy:

- During the intern phase, X Education should focus on leads with high predicted conversion probabilities (close to 1). Leads with favorable tags like "Will revert after reading the email" and positive profiles like "Potential Lead" should be prioritized for phone calls. By focusing on these leads, the interns can maximize conversions with minimal effort.

### 3. Post-Target Strategy:

- Once X Education meets its quarterly sales target, the sales team should shift focus to minimizing phone calls. A higher lead score threshold (e.g., 80% probability) should be used to determine which leads merit a phone call. Lower-probability leads should be contacted via less resource-intensive methods, like email or SMS, unless there is a strong business need to pursue them.

## Learnings

This assignment solidified my understanding of logistic regression as a predictive modeling technique, especially in binary classification tasks. I learned how to interpret logistic regression coefficients and translate them into actionable business insights. The exercise also underscored the importance of balancing technical accuracy with business objectives. For instance, while the model achieved an accuracy of 94%, the strategic recommendations were shaped by business considerations like resource allocation and lead prioritization.

Additionally, the importance of interpreting evaluation metrics like the ROC-AUC score (0.98) became clear, demonstrating that the model could effectively differentiate between potential conversions and non-conversions. Overall, this project provided valuable insights into how predictive models can drive business outcomes by improving decision-making and resource efficiency.