

# **Summary Report - X Education Lead Conversion Analysis**

## **Introduction:**

This summary report outlines the approach and findings in the analysis of X Education's lead conversion challenge. The goal was to increase lead conversion rates by identifying and prioritizing promising leads through a logistic regression model. The CEO set a target conversion rate of approximately 80%.

## **Approach:**

### **1. Data Preprocessing:**

- Data preprocessing was initiated, addressing issues such as missing values, treating categorical variables, and handling outliers.
- Categorical variables with a 'Select' level were treated as null values, as they represented a lack of information.

### **2. Feature Selection:**

- A total of 15 features were initially selected based on relevance.
- Feature selection was further refined by eliminating features with high p-values and high VIF (Variance Inflation Factor).

### **3. Data Splitting:**

- The dataset was divided into a training set and a test set for model development and evaluation.
- A logistic regression model was chosen for its interpretability and predictive capabilities.

### **4. Model Development:**

- A logistic regression model was trained on the training set.
- Features were selected based on their significance, as determined by p-values.
- Variance Inflation Factor (VIF) was used to address multicollinearity and eliminate redundant features.

### **5. Model Evaluation:**

- The model was evaluated on the test set, considering accuracy, sensitivity, and specificity metrics.
- Sensitivity and specificity played a pivotal role in optimizing the cutoff value for lead scoring.

### **6. Lead Scoring:**

- A lead score column was introduced to the test dataset to prioritize leads.
- Higher lead scores indicate leads are more likely to convert.

## **Key Findings:**

### 1. **Feature Selection:**

- A systematic approach was followed to select and eliminate features based on p-values and VIF.
- Features with p-values and VIF within acceptable thresholds were retained.

### 2. **Optimal Cutoff:**

- An optimal cutoff of 0.35 was chosen after rigorous evaluation, balancing accuracy, sensitivity, and specificity.
- Sensitivity (80.78%) closely aligned with the CEO's target conversion rate of around 80%.

### 3. **Model Efficiency:**

- The model demonstrated its efficiency with an accuracy of 80.27% on the test set.
- The sensitivity and specificity values further validated its effectiveness in identifying potential leads.

### **Conclusion:**

In conclusion, the analysis successfully addressed X Education's lead conversion challenge by employing a logistic regression model with 15 carefully selected features. Feature selection based on p-values and VIF ensured that only relevant variables were retained.

The model, with an optimal cutoff of 0.35, provides lead scores that effectively prioritize leads most likely to convert. Key features influencing conversion were identified, enabling the sales team to allocate resources more efficiently.

### **Recommendations:**

- Implement the logistic regression model into X Education's lead management process.
- Continuously monitor and update the model to adapt to evolving lead behavior.
- Provide training to the sales team on utilizing lead scores effectively for lead prioritization and communication.

This summary report highlights the journey from data preprocessing to model development and evaluation, emphasizing the importance of feature selection and its impact on the success of X Education's lead conversion strategy. The model's ability to identify and prioritize hot leads can significantly boost the company's conversion rates and overall success.