

# Lead Scoring Model for X Education: Summary Report

## Project Objective

The goal of this project was to develop a logistic regression-based lead scoring model to help X Education identify leads most likely to convert into paying customers. The model assigns a score between 0 and 100, allowing the sales team to prioritize efforts and improve the current conversion rate of 30% to 80%.

## Data Overview

The dataset consists of approximately 9,000 data points with features such as Lead Source, Total Time Spent on Website, Total Visits, and Last Activity, along with the target variable **Converted** (1 for converted leads, 0 for non-converted leads).

## Key Steps Performed

### 1. Data Cleaning:

- Removed columns with more than 30% missing values (e.g., Lead Profile).
- Treated missing values in relevant columns and standardized categorical variables (e.g., converting to lowercase).
- Replaced irrelevant categories (e.g., "Select") with NaN.

### 2. Exploratory Data Analysis (EDA):

- Visualized relationships between features and the target variable.
- Identified strong correlations between **Total Time Spent on Website** and lead conversion.
- Analyzed key categorical variables such as **Lead Source**, **Last Activity**, and **Lead Origin**.

### 3. Feature Engineering:

- Created dummy variables for categorical data (e.g., Lead Source).
- Standardized numerical variables like **Total Visits** to ensure proper scaling for modeling.

### 4. Model Building:

- Developed a logistic regression model using the cleaned dataset.
- Evaluated the model using accuracy, precision, recall, and ROC-AUC score to assess performance.

#### 5. Model Evaluation:

- Achieved high precision, successfully identifying converted leads.
- Key contributing variables include **Total Time Spent on Website**, **Lead Source**, and **Last Activity**.

### Insights and Recommendations

#### 1. Top Contributing Variables:

- **Total Time Spent on Website**: Leads who spend more time on the website have a higher chance of converting.
- **Lead Source**: Leads from **Referral** or **Google Search** are more likely to convert.
- **Last Activity**: Engagement actions like **Email Opened** or **Downloaded Brochure** significantly impact conversion probability.

#### 2. Focus on Categorical Variables:

- **Lead Source**: Prioritize marketing on platforms yielding higher conversion rates (e.g., Google, Referrals).
- **Last Activity**: Focus on leads showing recent, meaningful actions (e.g., opening emails or visiting key pages).
- **Lead Quality**: Implement a scoring system to segment leads based on their likelihood of conversion.

### Sales Phase Strategies

#### 1. Aggressive Conversion Phase (Interns Available):

- **Segmentation**: Classify leads into **Hot** (high conversion probability) and **Warm** (medium conversion probability) categories.
- **Resource Allocation**: Assign interns to handle Warm Leads through automated communication, while the sales team focuses on Hot Leads for personalized outreach.

- **Multichannel Approach:** Use phone calls, emails, and SMS to nurture high-priority leads efficiently.

## 2. Minimized Outreach Phase (Target Reached Early):

- **Raise Lead Score Threshold:** Only focus on the top 20% of leads with the highest scores.
- **Engagement-Based Filtering:** Prioritize leads who have exhibited recent engagement, such as visiting key pages or opening emails.
- **Automate Outreach:** Use email/SMS automation for low-priority leads, reserving phone calls for those with the highest potential.

## Business Impact

Implementing this lead scoring model will:

- **Streamline Sales Efforts:** Focus on high-conversion potential leads, ensuring efficient resource allocation.
- **Improve Conversion Rates:** Achieve the target conversion rate of 80%, significantly boosting the return on sales efforts.
- **Optimize Resources:** Adapt sales strategies during different phases (aggressive outreach vs. minimal outreach) to maximize resource efficiency.

## Next Steps

- **Model Deployment:** Integrate the lead scoring model into the sales pipeline for lead prioritization.
- **Strategy Implementation:** Execute the recommended outreach strategies based on the sales phase.
- **Ongoing Monitoring:** Continuously monitor and adjust the model to ensure it meets the desired conversion goals.

## Conclusion

This lead scoring model offers actionable insights for optimizing X Education's sales process. By targeting high-potential leads and adjusting strategies based on resource availability, the company can enhance its lead conversion rate and better allocate its sales resources.