Lead Scoring Case Study – Summary Report

Submitted by:
Nishant Anand: nishantax2024@email.iimcal.ac.in
Nishant Gaurav: nishantgaurav@gmail.com
Parul Chopra: parul6588@yahoo.co.in
Group: DS71

Problem Statement:
X Education is currently facing low lead conversion (~30%). Leads originate from Google, referrals, and social media, engaging via browsing, form submissions, and calls. The goal is to identify "Hot Leads" to improve efficiency and increase an 80% conversion rate.

Business Understanding & Data Overview:
Leads interact through various activities, and sales teams follow up via calls/emails. A logistic regression model assigns lead scores (0-100) to categorize leads. The dataset from 9000 leads approximately includes features such as Lead Source, Total Time Spent on Website, Last Activity, etc.

Following steps were performed in Python:

1. Import “Leads” Data
2. Inspect DataFrame
3. Data Preparation and EDA
4. Train-Test Split
5. Feature Scaling
6. Model Building
7. Feature Selection using RFE
8. Plotting the ROC Curve
9. Finding Optimal Cut-off Point
10. Precision and Recall
11. Making Predictions on the Test dataset
12. Lead Score Categorization

Data Processing & Analysis:

- Missing values were handled, categorical variables encoded, and numerical features were scaled.
- EDA was done. It revealed high engagement (more time spent) correlates with conversion.
- Feature selection was done using Recursive Feature Elimination (RFE) and Variance Inflation Factor (VIF) was used to select/drop features.

Model Development & Performance:

- Binomial Stats Model was used for Model build. 6 iterations were used to arrive at the final model.

- Logistic regression trained with a **ROC-AUC score of 0.88**.
- **Optimal cutoff probability: 0.33**, balancing sensitivity (81.02%) and specificity (79.13%).
- Test accuracy: **80.33%**, with **72.68% precision** and **80.45% recall**.

Lead Prioritization & Recommendations:

- **Hot & Warm Leads (60-100 score)** should be prioritized for direct sales.
- **Cold Leads (0-39 score)** to be engaged via automated emails/SMS.
- Enhance website engagement with chatbots and personalized content.
- Refine lead scoring dynamically, adjusting thresholds based on lead behaviour.
- Focused ad spend on high-performing sources (Google, direct traffic).

Conclusion:
The lead scoring model optimizes sales by prioritizing high-converting leads. AI-driven engagement and lead nurturing will further enhance conversion rates and business growth.