

Lead Scoring Case Study – Summary Report

Objective

X Education, an online course provider, wants to optimize its sales process by identifying high-potential leads. The goal of this project was to build a predictive model to assign a Lead Score to each lead, indicating the probability of conversion. The company has an 80% target lead conversion rate, in which case there needs to be a strategic concentration on the most promising leads.

Approach & Execution

1. Data Exploration & Preprocessing:

We started by loading and understanding the dataset, which contained 9,240 leads with 37 features. The dataset involved a combination of numerical and categorical fields like Lead Source, Last Activity, and Total Time Spent on Website. Key steps in preprocessing were:

- Handling Missing Values: Columns with excessive missing values were dropped, while others were imputed using the mode.
- Encoding Categorical Variables: Converted categorical features into numerical representations.
- Feature Scaling: Standardized numerical variables to bring them onto the same scale.
- Splitting Data: Divided the dataset into training (70%) and testing (30%) sets.

2. Model Building:

We selected Logistic Regression, which is a widely used classification model for binary outcomes, because it is easy to interpret and it is also quite efficient. The model was trained to predict whether a lead would convert into a paying customer (1) or not (0). Decision Trees model was considered but it was overfitted.

Key insights from model performance:

- Accuracy: ~91% on test data
- ROC-AUC Score: 0.96 (which means it's a strong predictive model)
- Optimal Threshold: 0.42 (this was so that we can balance precision and recall)

3. Key Findings – Important Features:

The top three numerical features contributing to lead conversion are the following:

- Total Time Spent on Website – More time spent means higher the interest.
- Lead Origin – Leads from "API" and "Landing Page Submission" converted more.
- Last Activity – Engagement through emails/SMS increased chances of conversion.

The top categorical features that should be prioritized for higher conversion rates are:

- Lead Source – Organic Search, Direct Traffic, and Google performed well.
- Lead Quality – "High" and "Might Be" had better conversion chances.
- Specialization – Some fields, like Finance and Banking, had higher conversion.

Strategic Recommendations:

For Aggressive Lead Conversion During Intern Hiring Period

- Lowering the Lead Score Threshold (like from 50 to 30) would increase the pool of leads.
- Prioritizing High-Engagement Leads (those who interacted with emails, watched videos, or revisited the site).
- Applying a Multi-Touch Approach – Combining emails, SMS, and calls to improve the response rates.

For Reducing Unnecessary Calls After Achieving the Targets

- Increasing the Lead Score Threshold (like focusing only on leads scoring 80+).
- Prioritizing Referral Leads, as they have a higher chances of conversion.
- Automating Follow-ups via emails and SMS before assigning a sales call.

Key takeaways:

- Data Quality Matters: Missing values and inconsistent categories require careful preprocessing.
- Feature Importance Insights: User behavior (time spent, last activity) they have a significant impact on lead conversion.
- Threshold Selection is Crucial: Adjusting the probability threshold will help in balancing the business needs (more calls vs fewer, high-quality leads).
- Model Simplicity vs Interpretability: Logistic regression worked well due to its clarity and strong results, making it preferable over other complex models.
- By implementing this model, X Education can improve their efficiency by focusing on high-scoring leads, potentially increasing conversions by 10-20%.

This project gave useful insights into lead prioritization, helping businesses increase their efficiency and focus efforts on high-converting leads.