

Lead Scoring Case Study

Jibin
Shreya
Prashanth

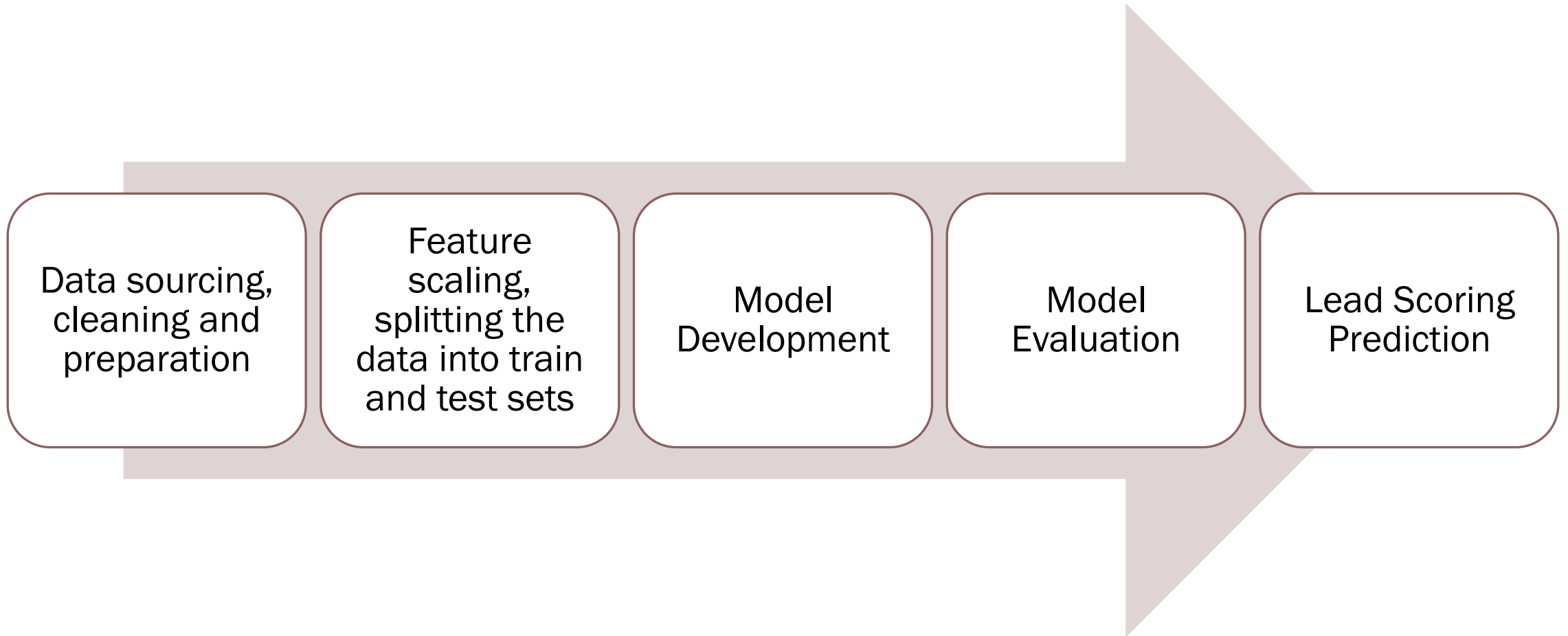
Problem Statement

X Education, an online education company, has a significant challenge with low lead conversion rates. Despite acquiring a large number of leads (via marketing efforts, website visits, and referrals), only 30% of leads are converted into paying customers. The company aims to improve this conversion rate by identifying the "Hot Leads" — those that are most likely to convert — and prioritizing them for follow-up by the sales team. This approach will help optimize resources, increase efficiency, and ultimately raise the lead conversion rate to approximately 80%.

Objective

The goal is to build a predictive model that scores leads based on their likelihood of conversion, using historical data on lead activities, behavior on the website, and other relevant factors. The leads with higher scores should have a higher probability of conversion, allowing the sales team to focus on the most promising leads.

Analysis Methodology



EDA

- Handled Null values.

- Dropped columns with more than 40% null values.

- Handled Categorical Values by using get dummies method.

Variables impacting the Conversion Rate

- TotalVisits
- Total Time Spent on Website
- Lead Source_Welingak Website
- Last Activity_SMS Sent
- Country_Germany
- Tags_Closed by Horizzon
- Tags_Lost to EINS
- Tags_Ringing
- Tags_Will revert after reading the email
- Tags_switched off

Model Evaluation – Train Set

Confusion Matrix :

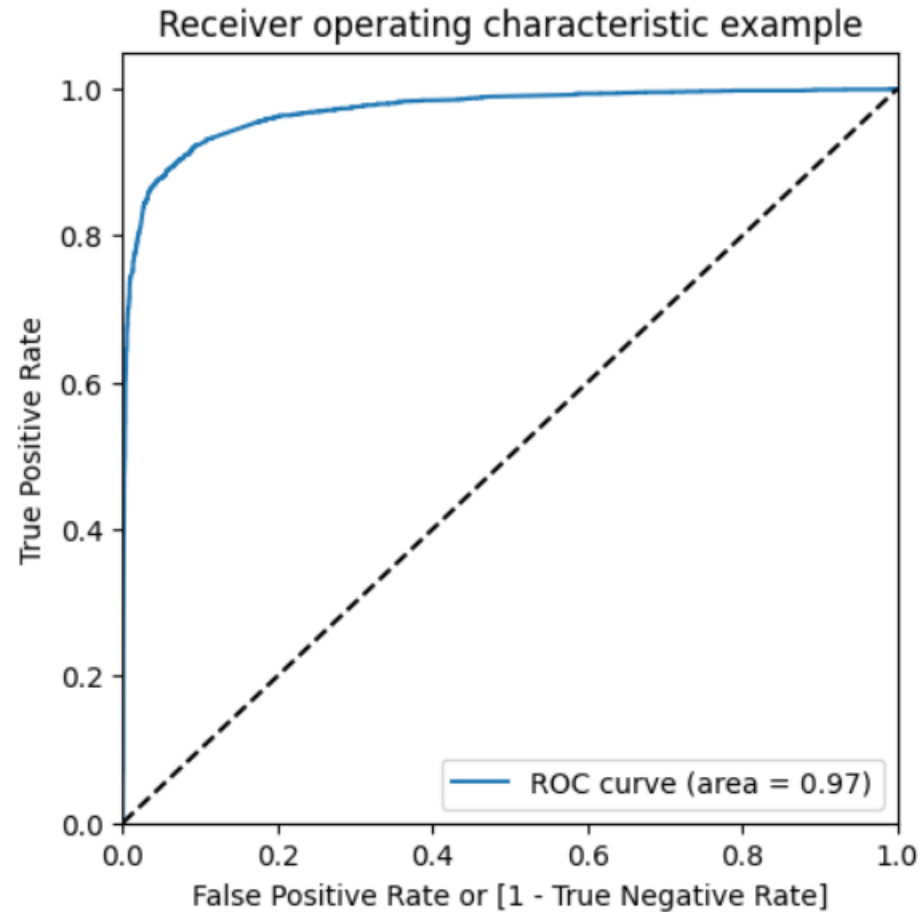
| Actual/Predicted | Negative | Positive |
|------------------|-----------------------|-----------------------|
| Negative | True Negative 3763 | False Positive 152 |
| Positive | False Negative 323 | True Positive 2113 |

Model Performance:

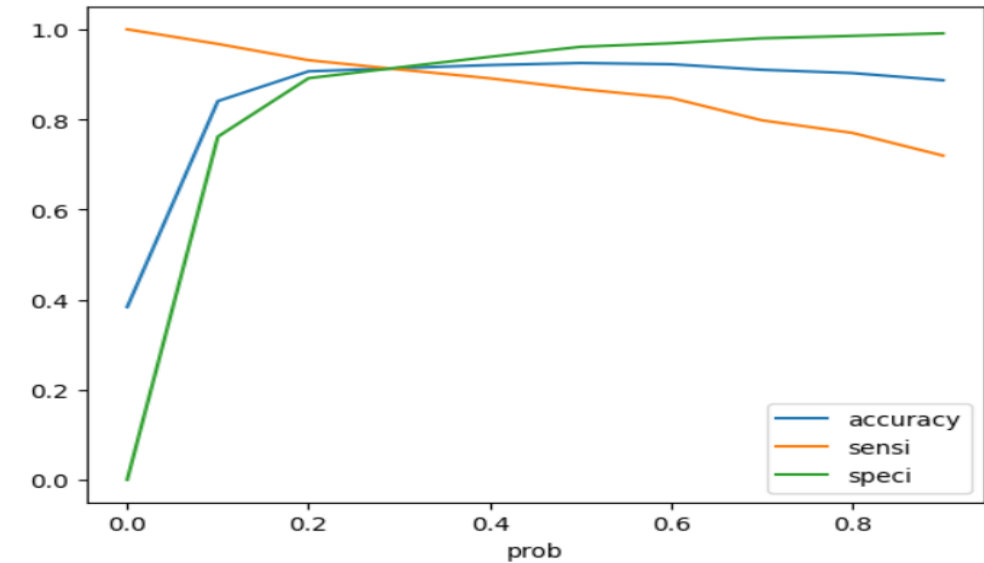
| | |
|---------------------------|--------|
| Accuracy | 91.3 % |
| Sensitivity | 91.0% |
| Specificity | 91.5% |
| False Positive | 8.45% |
| Positive predictive value | 87% |
| Negative predictive value | 94.2% |

Model Evaluation – Train Set

ROC Curve



Model Performance:



From the curve above, 0.3 is the optimum point to take it as a cutoff probability.

| | |
|-----------|--------|
| Precision | 93.2 % |
| Recall | 86.7% |

Model Evaluation – Test Set

Confusion Matrix :

| Actual/Predicted | Negative | Positive |
|------------------|-----------------------|-----------------------|
| Negative | True Negative 1569 | False Positive 155 |
| Positive | False Negative 93 | True Positive 906 |

Model Performance:

| | |
|---------------------------|--------|
| Accuracy | 92.5 % |
| Sensitivity | 85.4% |
| Specificity | 91% |
| False Positive | 8.9% |
| Positive predictive value | 90.6% |
| Negative predictive value | 94.4% |

Train and Test Data Comparison :

| | Train Data | Test Data |
|---------------------------|------------|-----------|
| Accuracy | 91.3 % | 92.5 % |
| Sensitivity | 91.0% | 85.4% |
| Specificity | 91.5% | 91% |
| False Positive | 8.45% | 8.9% |
| Positive predictive value | 87% | 90.6% |
| Negative predictive value | 94.2% | 94.4% |

The model metrics for test and train data are very similar and higher than 85%

=> Good Performance of the model