

# Lead Scoring Case Study

# Problem Statement

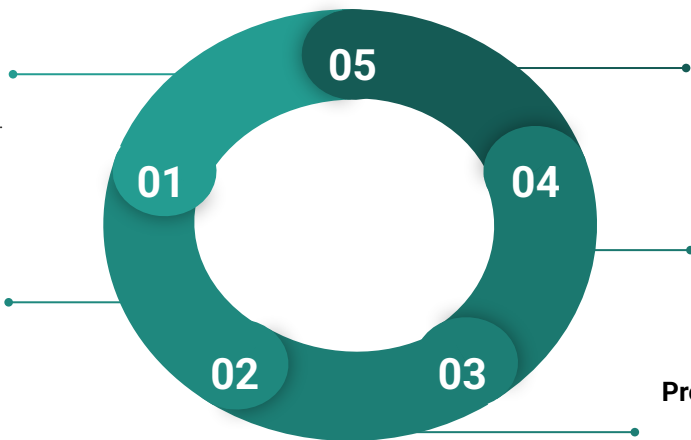
Identify the set of leads of X Education so that the lead conversion rate should go up and the sales team of the company focus more on communication with the potential leads rather than making calls to every customer.

## Data understanding and Exploring

- Understand data distribution.
- Visualize and identify key insights.

## Data Cleaning

- Handle missing values.
- Treat outliers.



## Test the model

- Test the final model on the test dataset.
- Measure performance metrics (accuracy, sensitivity, specificity).
- Compare test set results with training set performance.

## Train the Model

- Train initial model using all features.
- Optimize features (REF, VIF, p-values).
- Predict customer conversion with an optimal probability cut-off.
- Evaluate model performance: Accuracy, Sensitivity, Specificity, Gini coefficient

## Prepare the data for modeling

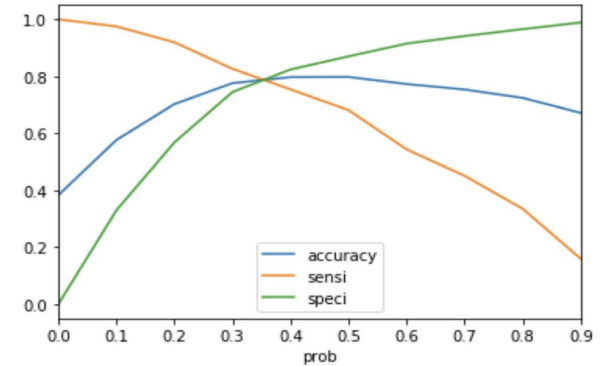
- Address data imbalance.
- Convert categorical variables: Binary encoding and dummy variables.
- Split data into train and test sets.
- Scale the data for uniformity.

# Model Features

| Features                                | Coefficient | Impacts (On decreasing order) |
|---|-------------|-------------------------------|
| Lead Source_Welingak Website            | 4.5256      | Positive                      |
| Lead Source_Reference                   | 3.1268      | Positive                      |
| Last Activity_Olark Chat Conversation   | -2.8298     | Negative                      |
| Last Activity_Converted to Lead         | -2.4484     | Negative                      |
| Last Activity_Form Submitted on Website | -1.9713     | Negative                      |
| Last Activity_Email Bounced             | -1.8597     | Negative                      |
| Last Activity_Email Link Clicked        | -1.6500     | Negative                      |
| Last Activity_Page Visited on Website   | -1.5718     | Negative                      |
| Do Not Email                            | -1.3403     | Negative                      |
| Lead Source_Direct Traffic              | -1.3241     | Negative                      |
| Lead Source_Referral Sites              | -1.2794     | Negative                      |
| Total Time Spent on Website             | 1.1267      | Positive                      |
| Last Activity_Email Opened              | -1.1030     | Negative                      |
| Lead Source_Organic Search              | -1.0773     | Negative                      |
| Lead Source_Google                      | -0.9545     | Negative                      |

# Evaluation Matrix

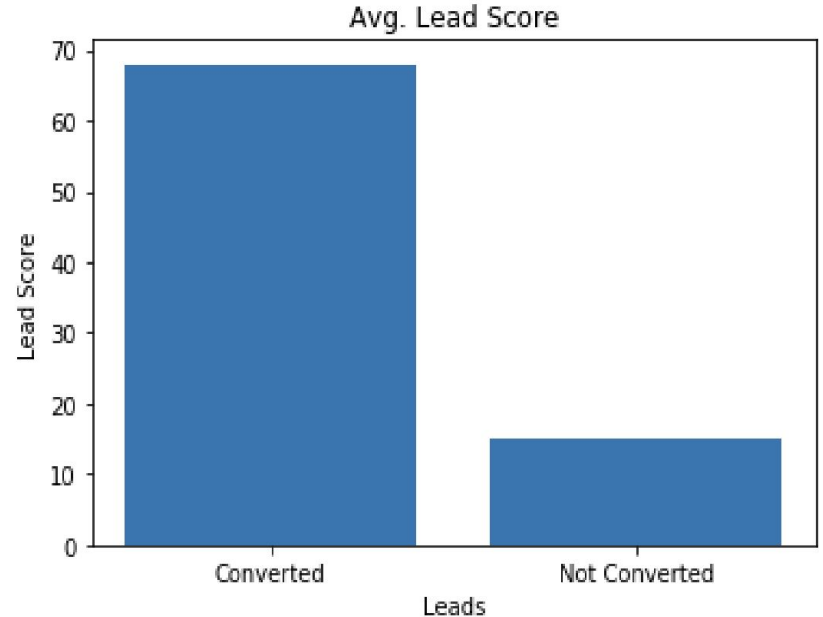
| Measures  | Train set | Test set |
|---|-----------|----------|
| Accuracy  | 0.79      | 0.79     |
| Sensitivity<br>(No. of correctly predicted YES) | 0.79      | 0.79     |
| Specificity<br>(No. of correctly predicted NOs) | 0.78      | 0.78     |
| Gini (ROC curve area)                           | 0.86      |          |



- The optimal predicted probability for a customer to be converted is 0.35.
- The model demonstrates strong performance with good accuracy, sensitivity, and specificity.
- It achieves a balanced trade-off among accuracy, sensitivity, and specificity, ensuring reliable predictions.

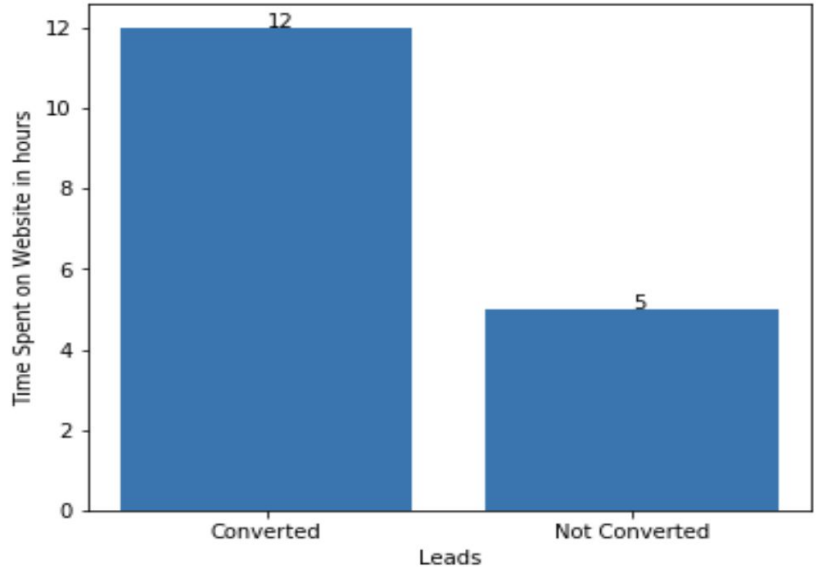
# Lead Score

- The Lead Score, ranging from 0 to 100, reflects the probability of a customer converting.
- A higher Lead Score indicates a greater likelihood of conversion.
- Customers expected to convert have an average Lead Score of 68.
- Customers not expected to convert have an average Lead Score of 15.



# Average Total Time Spent on Website

- Converted customers spend an average of 12 hours on the website.
- Non-converted customers spend an average of 5 hours on the website.



# Business recommendation for higher conversion rate

## Highly Likely to be Converted Leads:

1. Lead Score exceeds 68.
2. Total time spent on the website is over 12 hours.
3. Lead sources are either Welingak Website or Reference.

## Very Less Likely to be Converted Leads:

1. Customers who opted for the "Do Not Email" option.
2. Lead Score is below 15.
3. Total time spent on the website is under 5 hours.
4. Lead sources include Direct Traffic, Referral Sites, Organic Search, or Google.
5. Last recorded activity includes any of the following:
  - Olark chat conversation
  - Page visited on the website
  - Email bounced
  - Form submitted on the website
  - Email link clicked

Thank You!