# Lead Scoring Case Study
# X - Education

Optimizing Lead Conversion Analysis
Logistic Regression Model

# Problem Statement

## Current Challenges

X Education struggles with a low conversion rate, converting only about 30% of its leads into customers.

## Our Goal

We aim to develop a predictive model to assign lead scores, targeting an ambitious conversion rate of 80%.

## Actionable Strategies

We will identify and prioritize high-potential leads through targeted communication and tailored follow-up strategies.

# Project objective:

To develop a model that accurately scores leads, enhancing conversion rates for X Education

# Data Overview

- The dataset includes 9,240 leads with features such as Lead Source, Total Time Spent on Website, Tags, Last Activity, etc.
- The target variable is 'Converted', indicating if a lead was converted (1) or not (0).

# Key features:

**Lead Source**: The origin of the lead (e.g., Google, referral sites) plays a crucial role in conversion probability.

**Total Time Spent on Website**: Indicates the level of interest; leads spending more time are often more engaged.

**Last Activity**: Tracks the last interaction type, such as email, phone call, or website visit, providing insight into lead engagement.

**Do Not Email Flag**: Shows whether a lead prefers not to be contacted via email, affecting follow-up strategies.

**Current Occupation**: Leads who are working professionals tend to have a higher likelihood of conversion based on past data.

# Approach and Methodology

- Data Understanding & Exploration
- Data Cleaning Approach
- Feature Engineering
- Data Splitting and Scaling
- Correlation Analysis
- Model Selection
- Key Model Insights
- Model Evaluation Metrics
- Optimal Probability Cutoff
- Test Set Predictions and Validation

# Data Understanding & Exploration

- Dataset Overview: 9,240 data points with attributes related to lead interactions and demographics.
- Key Features Explored: Lead Source, Total Time Spent on Website, Last Activity, etc.

# Data Cleaning Approach

- Removed columns with over 3000 missing values.
- Dropped rows with uninformative values like "Select."
- Eliminated features with predominantly one-sided values (e.g., 95% “No”).

# Feature Engineering

- Dummy Variables: Created for categorical features to enrich model inputs.

- Standardization: Applied scaling for numerical variables to improve model stability.

# Data Splitting and Scaling

- Test-Train Split: Divided data into training and test sets.

- Scaling: Standardized continuous variables for consistency across features.

# Correlation Analysis

- Objective: To identify feature relationships and avoid multicollinearity.

- Outcome: Features selected with low VIF and relevant correlations.

# Model Selection

- Model Type: Logistic Regression with Generalized Linear Model (GLM).

- Selection Criteria: High interpretability and suitability for binary classification.

# Model Results

- Key Model Metrics: Pseudo R-squared of 0.38, indicating a moderate fit.

- Significant Variables: Lead Origin, Time Spent on Website, Last Activity, Occupation.

# Key Model Insights

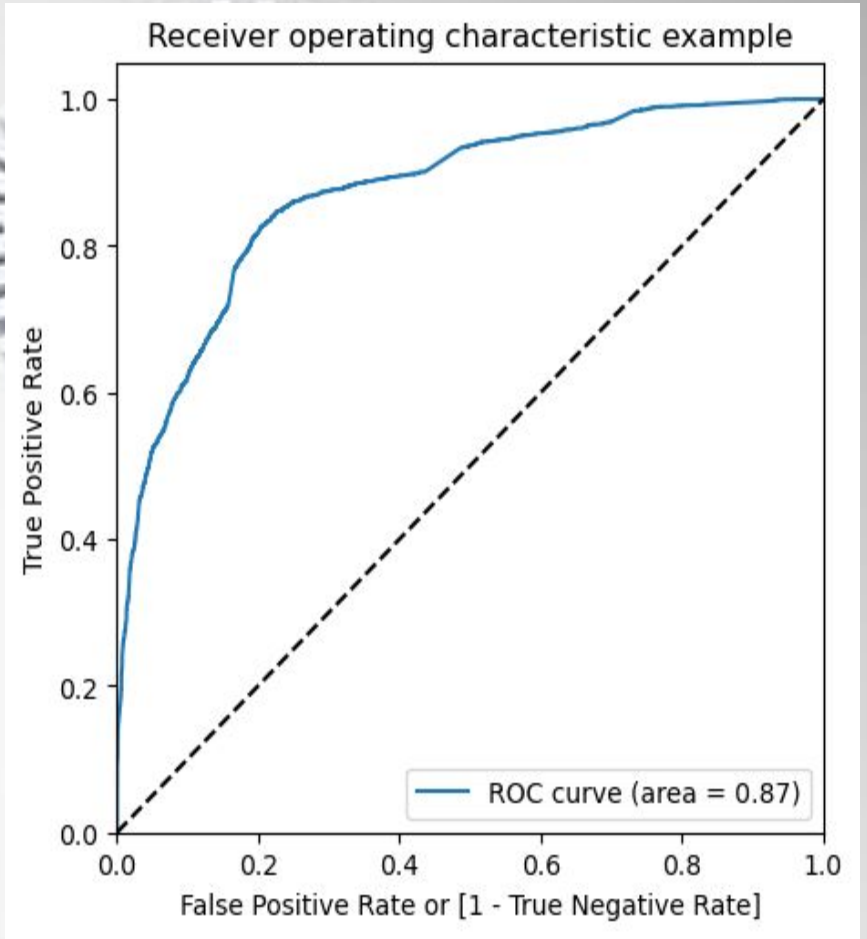
Top Predictors:

- Time Spent on Website: Strong positive indicator.
- Lead Origin: Higher likelihood from specific forms.
- Last Activity & Current Occupation: Valuable indicators of conversion likelihood.

# Model Evaluation Metrics

- Evaluation: Confusion Matrix, Sensitivity, and Specificity analysis.

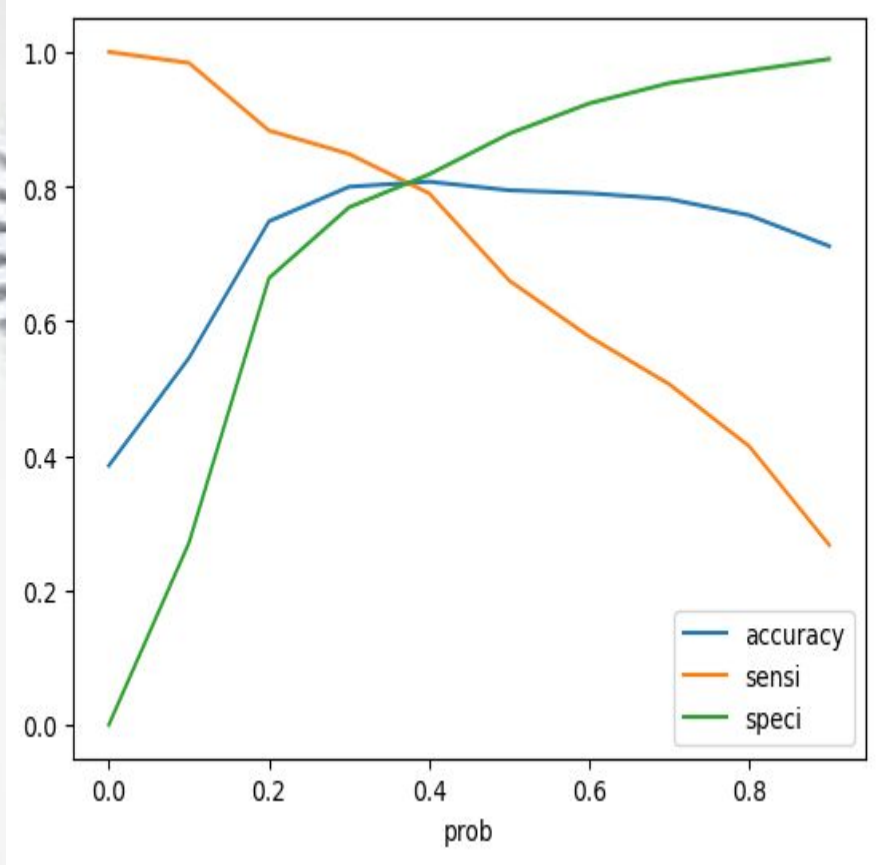
- ROC Analysis: AUC score of 0.87, showing strong model performance.



# Optimal Probability Cutoff

- Cutoff Optimization: Selected 0.40 for optimal sensitivity, specificity, and accuracy.

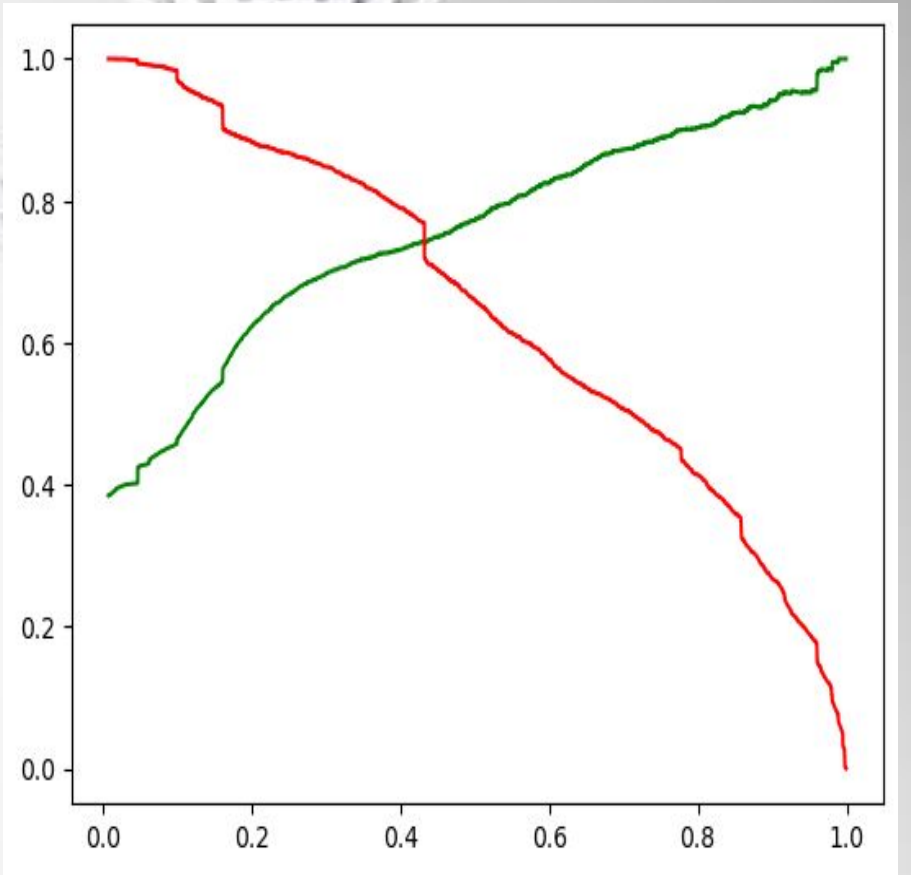
- Result: Balanced metrics for real-world application.



# Test Set Predictions and Validation

- Test Accuracy: Close alignment with training results.

- Precision-Recall Trade-off: Ensures reliable performance on unseen data.



# Insights:

- Leads spending more time on the website have a higher conversion likelihood.
- Lead Add Form origin strongly correlates with conversions.
- Working professionals are more likely to convert.
- Phone conversations as the last activity drive conversions.
- “Do Not Email” preference negatively impacts conversion rates.

# Recommendations:

- Prioritize follow-ups for leads with high website engagement.
- Focus resources on Lead Add Form and similar channels.
- Target marketing efforts toward working professionals.
- Allocate sales resources to leads with impactful activities like phone calls.
- Use alternative channels for leads who opted out of email.

# CONCLUSION

In conclusion, our analysis highlights the importance of identifying and nurturing high-potential leads to enhance conversion rates. By leveraging insights on lead behavior and preferences, X Education can optimize its sales strategies, ultimately aiming for a lead conversion rate that aligns with the company's targets. Implementing the recommended strategies will not only streamline the lead conversion process but also foster stronger relationships with potential customers, contributing to the overall growth and success of the organization.