# Lead Conversion Analysis

# Problem Statement

- X Education sells online courses but faces a low lead conversion rate (around 30%).
- The goal is to identify high-potential leads (Hot Leads) to increase the conversion rate to approximately 80%.
- Importance of optimizing lead conversion for business growth.

# Lead Generation Process

Involves three stages:

**Top Stage: Many** leads generated.

**Middle Stage:** Need for nurturing potential leads through:
Educating about products.
Constant communication.

**Bottom Stage:** Few leads convert to paying customers.

**Our objective here is to nurture the middle stage**

# Data Overview

**Dataset:**

- Size: Approximately 9,000 leads.
- Key Variables: Lead Source, Total Time Spent on Website, Total Visits, Last Activity,
- Converted (target variable).

- **Data Challenges:**
- Missing values and categorical variables needing transformation.

# Data Cleaning Process

**Steps Taken:**

- Created dummy variables for categorical features.
- Removed columns with >3,000 missing values.
- Treated 'Select' values as missing.
- Dropped columns with predominantly one value.
- Retained 69% of data after handling missing values.

# Feature Engineering & Model Building

**Feature Selection:**

- Used Recursive Feature Elimination (RFE) to identify 15 key variables.

**Model Development:**

- Logistic regression model created using statsmodels.
- Ensured p-values < 0.05 and assessed VIF for multicollinearity.
- Eliminated variables with high p value>0.05 and high VIF values >5

- **Selected Features:**

Lead Source_Welingak Website

Do Not Email_Yes

Last Activity_SMS Sent

What is your current occupation_Student

What is your current occupation_Unemployed

Last Notable Activity_Modified

Last Notable Activity_Unreachable

# Model Results

**Performance Metrics:**

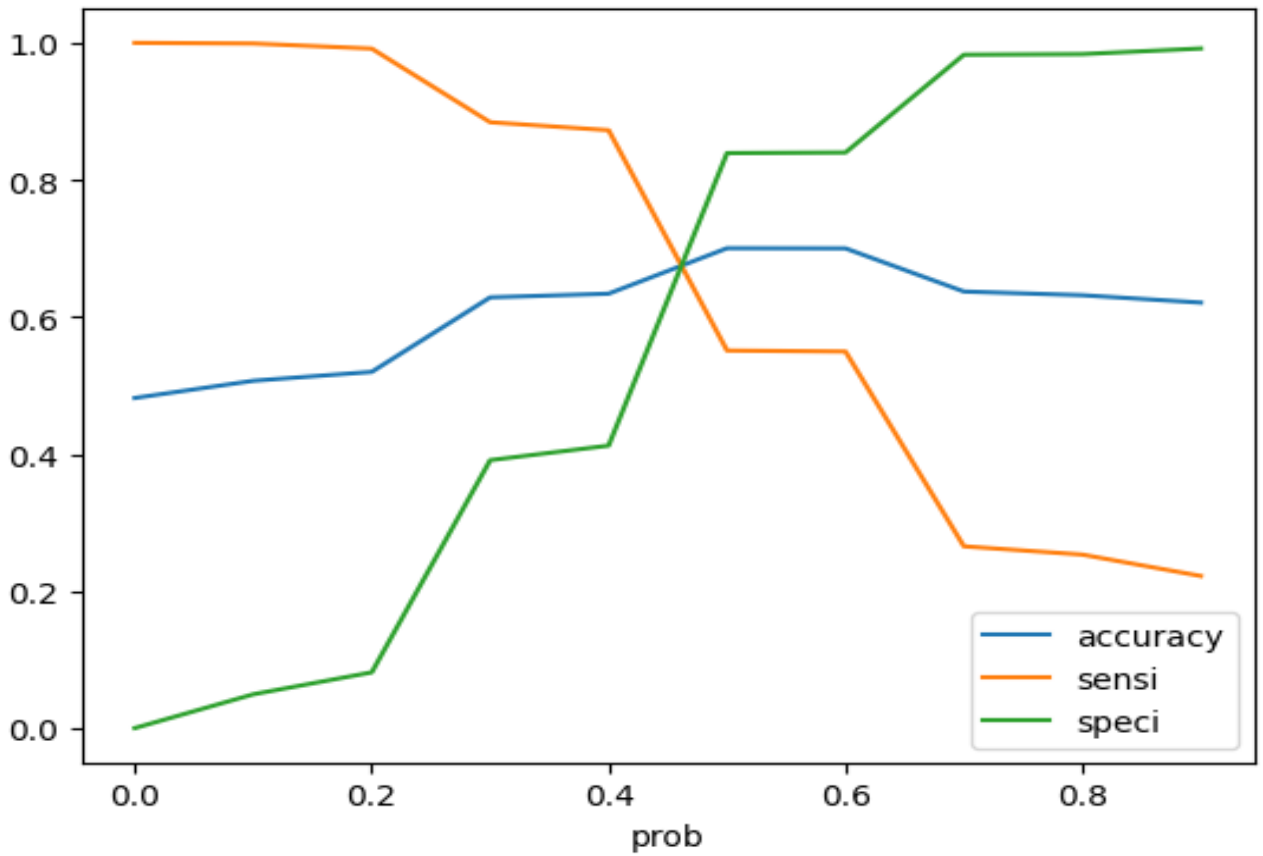
- Accuracy: ~70%
- Precision: 71%
- Recall: 62%

**ROC Analysis:**

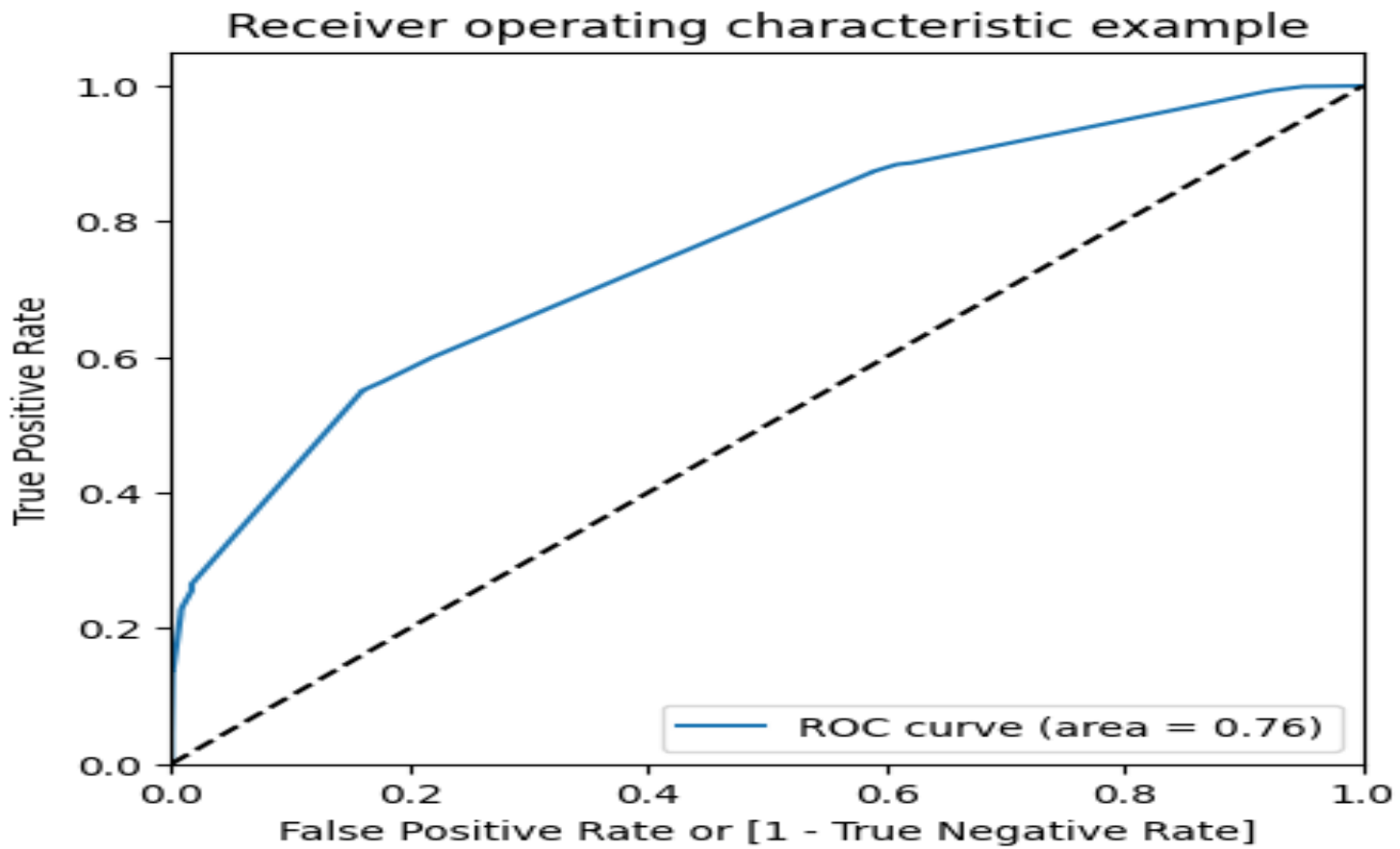
- AUC: 0.76, indicating good model performance.
- Optimized cutoff at 0.42 for better sensitivity and specificity.

# MODEL RESULTS

- **OPTUM CUTOFF**



# ROC


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.76)

# FEATURE IMPORTANCE

| Rank | Feature | Coefficient |
|---|---|---|
| 1 | Lead Source_Welingak Website | 4.712322 |
| 2 | What is your current occupation_Unemployed | -2.755510 |
| 3 | What is your current occupation_Student | -2.556026 |
| 4 | const | 2.389669 |
| 5 | Last Notable Activity_Unreachable | 2.090588 |
| 6 | Do Not Email_Yes | -1.477350 |

# Performance Metrics Interpretation

**Accuracy: ~69.6%**

- Approximately 70% of predicted converting leads are accurate, enabling the sales team to focus efforts effectively.

**Precision: 71%**

- When the model predicts a lead will convert, it is correct 71% of the time, reducing wasted efforts on low-potential leads.

**Recall: 61%**

- The model identifies 62% of actual conversions, indicating potential to improve lead capture and nurturing strategies.

**ROC Analysis: AUC: 0.76**

- Good discriminatory power to distinguish between converting and non-converting leads, enhancing targeting efficiency.

**Optimized Cutoff at 0.42**

- Classifies leads as likely to convert if their probability exceeds 0.42, balancing sensitivity and specificity for strategic lead targeting.

# Key Findings

- **Top Predictors of Conversion:**
  - Lead Source
  - Last Occupation
  - Last Notable Activity
- **Critical Categorical Variables:**
  - lead_source_welingnak
  - Current Occupation: Unemployed
  - Current Occupation: Student

# Recommended Strategies

**For Aggressive Outreach:**

- Focus on high-impact variables and categorical segments.
- Concentrate on "welingnak" leads, unemployed, and students.

**For Focus on New Work:**

- Prioritize leads with high likelihood of conversion only when necessary.
- Minimize outreach to reduce ineffective calls.

# THANK YOU