Lead Score Case Study

Submitted By:

Anuj Tiwari

Vikas Tunwal

Trang Van

Lead Score Case Study in X Education

Problem Statement :
Company X Education is selling online courses to industry professionals. The company markets its courses on several websites and search engines like Google.
Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals.
Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

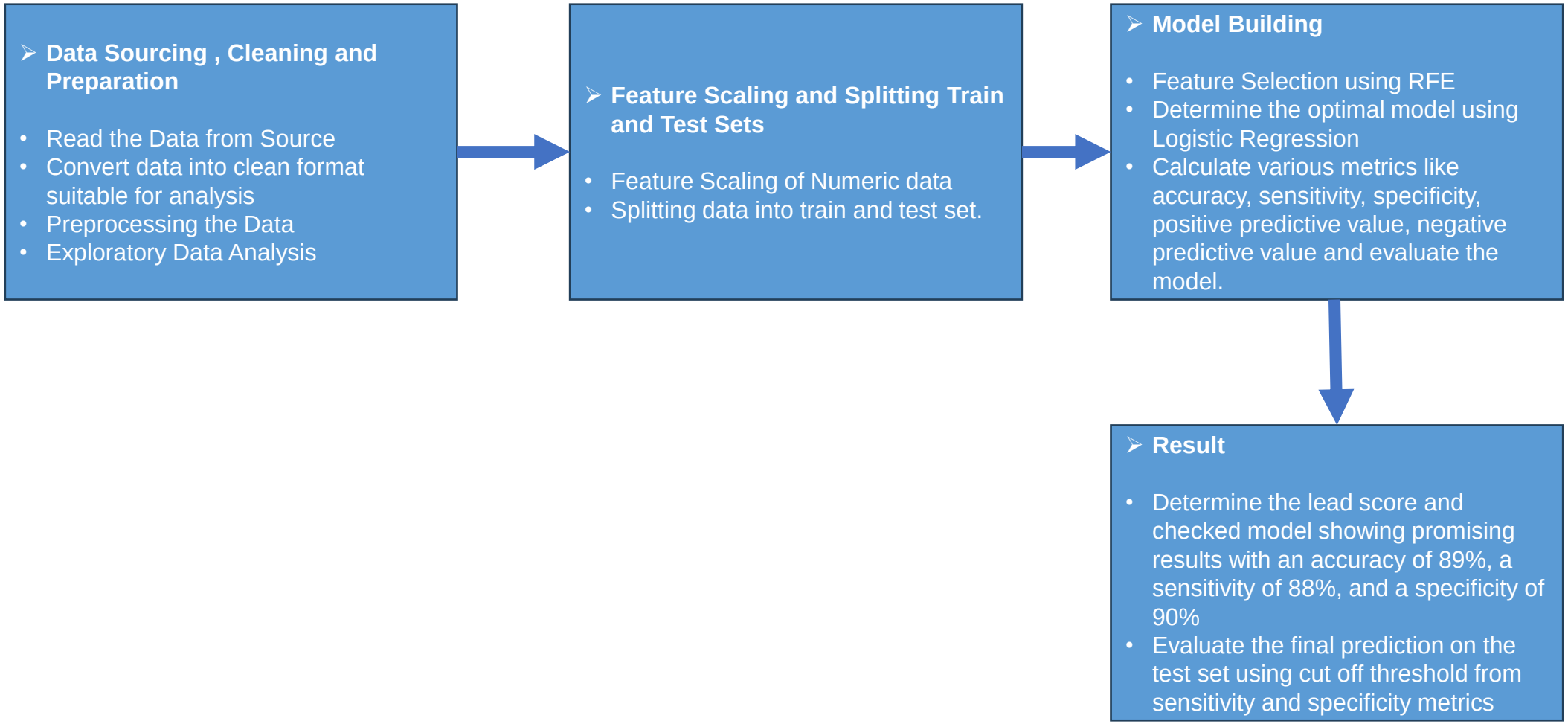
Objective:

X Education wants to know the most promising leads, i.e. the leads that are most likely to convert into paying customers.
The company needs a model where we assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. Desired lead score ranges from 0 to 100.
The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Strategy

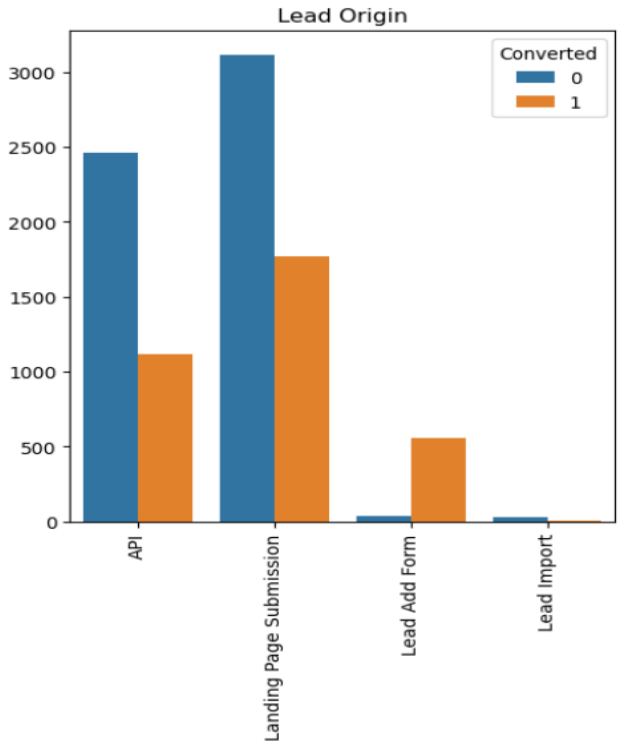
- Importing the required libraries and data Ingesion for analysis.
- Cleaning and Understanding the Data.
- Data Preprocessing and Exploratory Data Analysis.
  - Univariate
  - Bivariate
- Data Splitting.
- Scaling the Features.
- Building logistic Regression models and Refinement.
- Evaluating the model by metrics -Specificity and Sensitivity or Precision and Recall.
- Plotting ROC Curve.
- Determining the Optimal Cut-off Point.
- Applying the best model in Test data to get the predictions and Model Validation.
- Final Model Metrics.

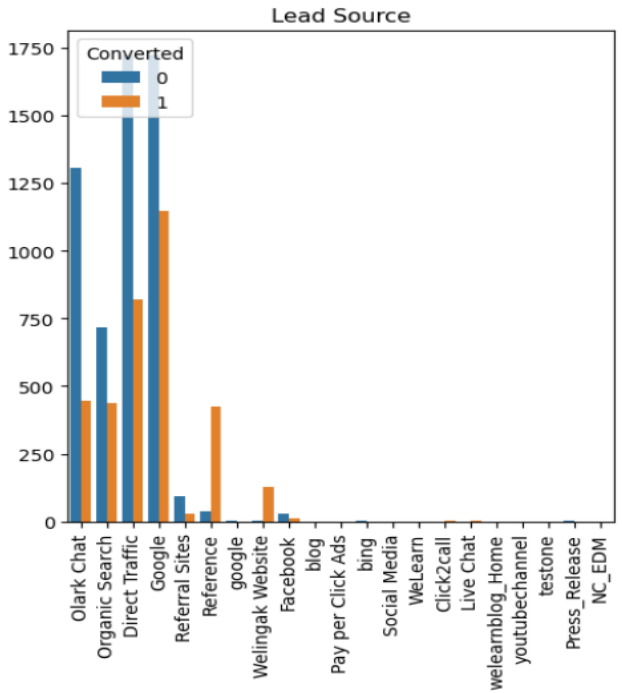
Problem solving methodology

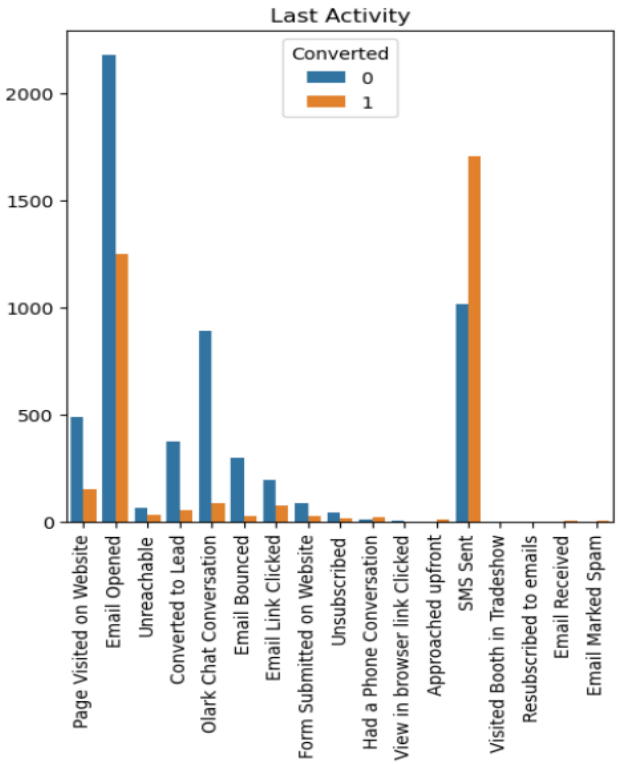


Exploratory Data Analysis

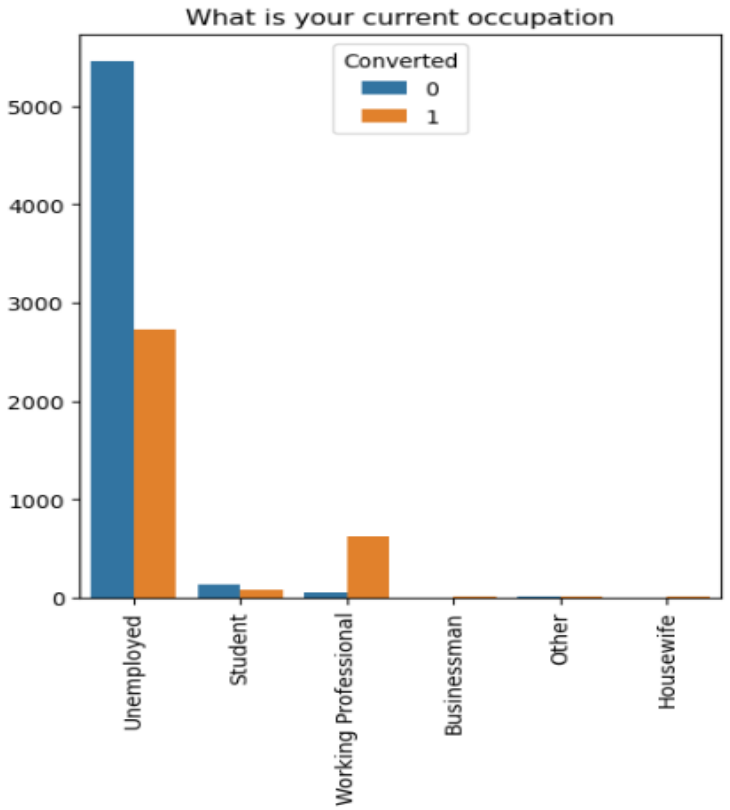
- Maximum conversions in Lead Origin happened from Landing page submission, Lead Source from Google, Last Activity from SMS Sent

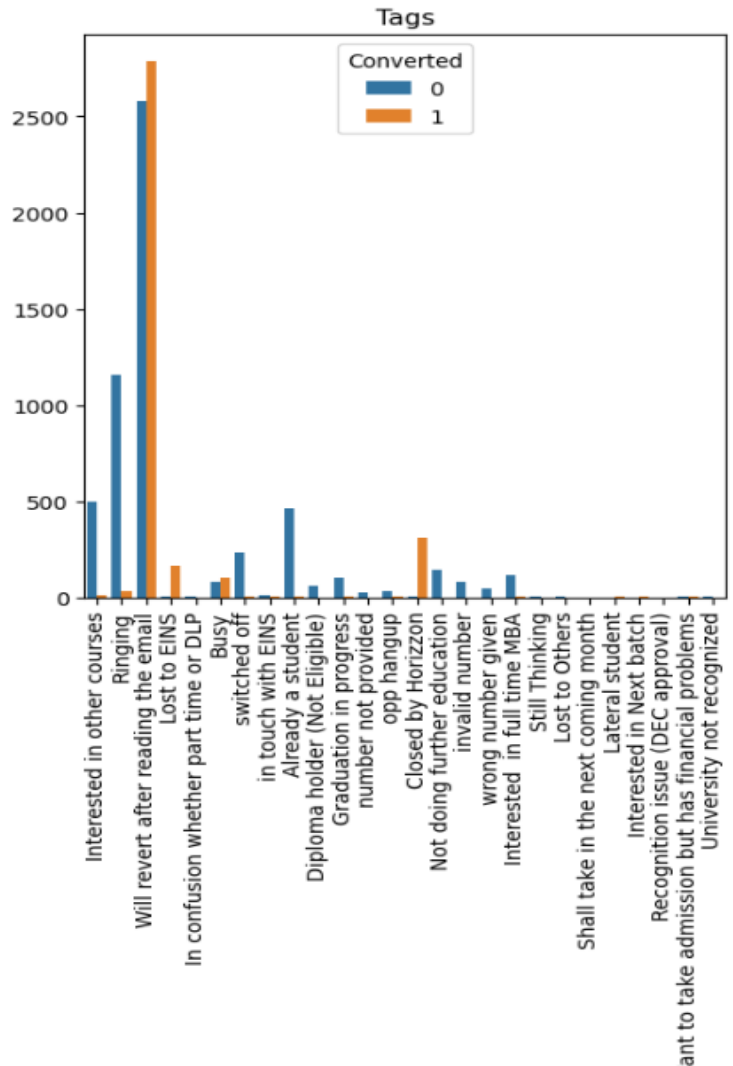




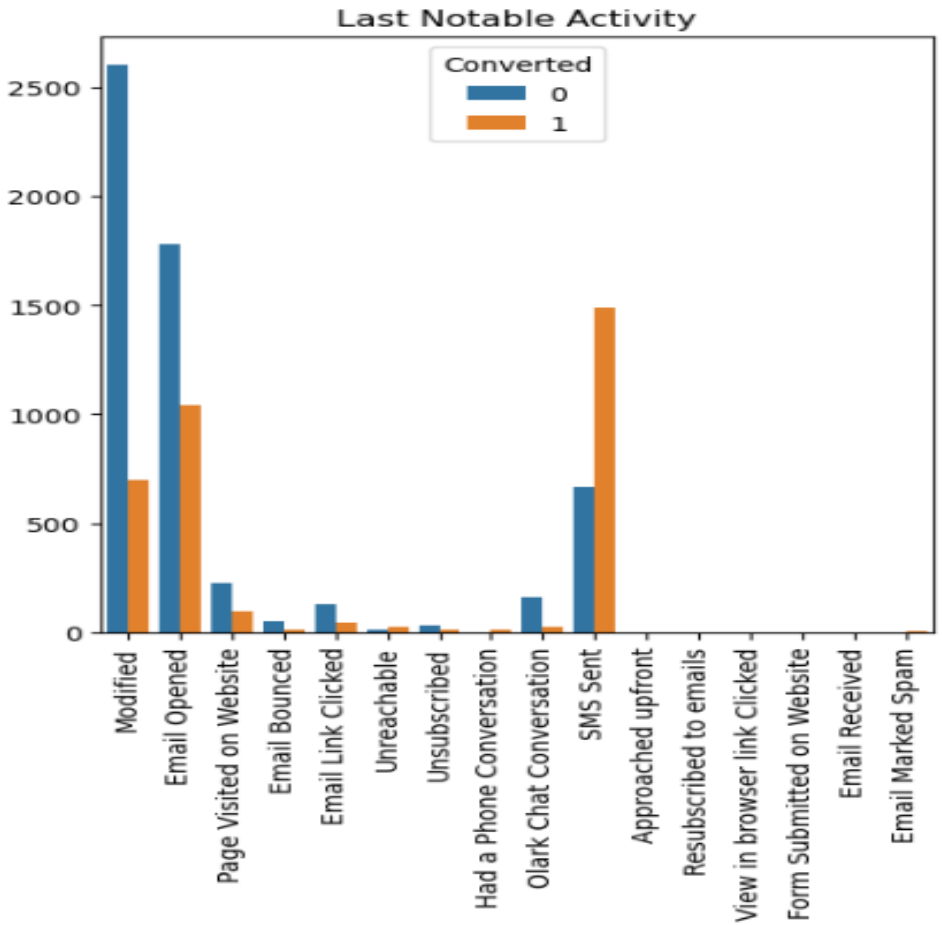


Most of the conversions happened in Current Occupation from “Unemployed”, in Tags from “Revert after reading the email”





Major conversion in Last Notable Activity from “SMS Sent” and “Email Opened”



Features which indicates higher likelihood of lead conversion

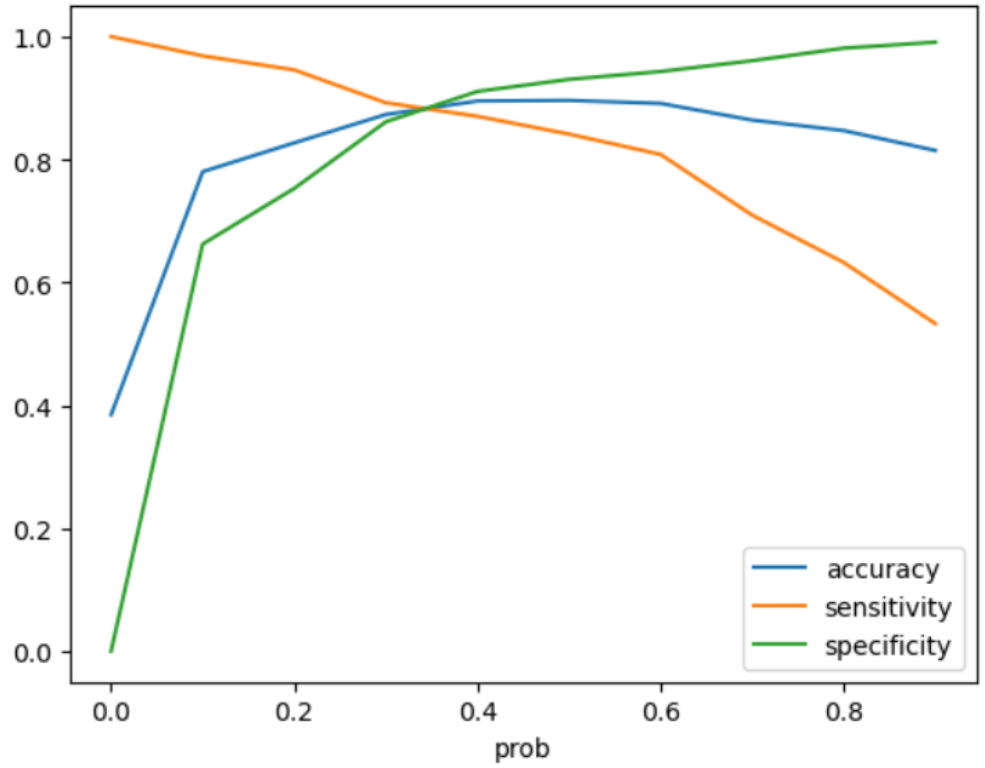
- Tags_Closed by Horizon
- Tags_Lost to EINS
- Tags_Will revert after reading the email
- Tags_Busy
- Lead Origin_Lead Add Form

Features which indicates lower chances of lead conversion

- Tags_switched off
- Last Notable Activity_Olark Chat Conversation

Optimal Cut Off point

According to the plot Optimal Cut Off point is 0.34



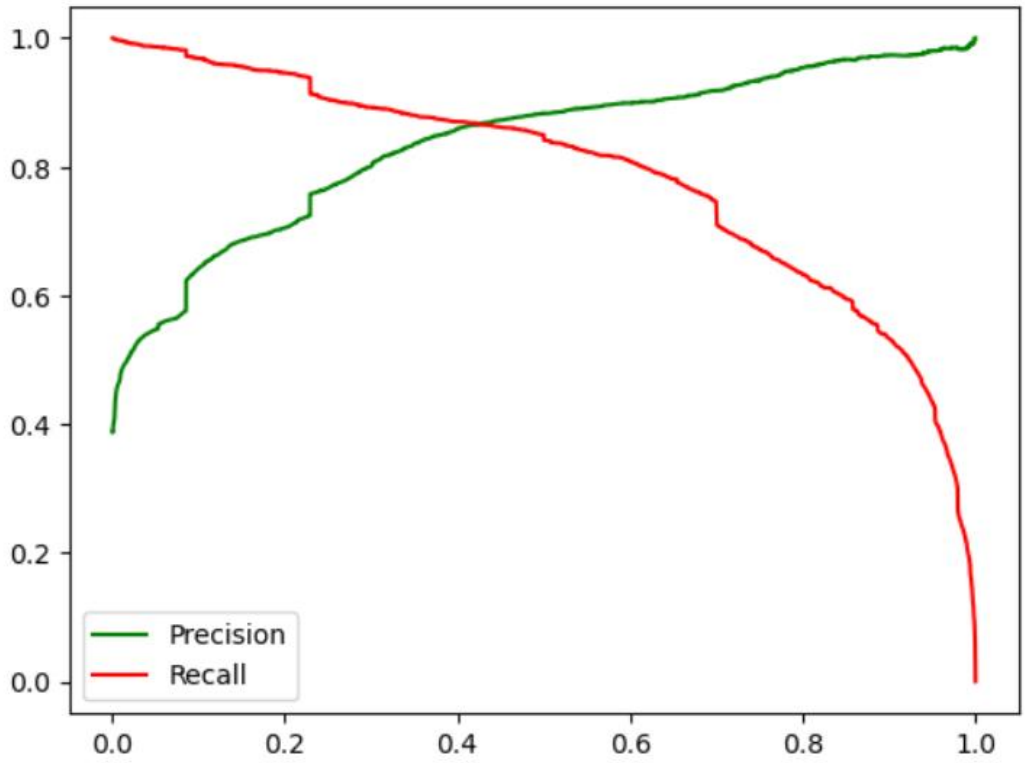
Sensitivity, Specificity, Precision, Recall

- Overall Model Accuracy-88.74%
- Sensitivity-88%
- Specificity-89.21%
- Positive predicted value-83.60%
- Negative predicted value-92.24%
- Precision-88.30%
- Recall-84.12%

Confusion Matrix

```
array([[3498, 423],
       [ 294, 2157]]
```

Trade-off between Precision and Recall



Final Summary: Key Factors Influencing Lead Conversion

We've identified several major contributors to lead conversion based on our analysis. Here they are, along with their respective coefficients:

- **Tags_Closed by Horizon:** Coefficient - 8.9448
- **Tags_Lost to EINS:** Coefficient - 8.5327
- **Tags_Will revert after reading the email:** Coefficient - 3.4580
- **Tags_Busy:** Coefficient - 3.1989
- **Lead Origin_Lead Add Form:** Coefficient - 3.0088
- **What is your current occupation_Working Professional:** Coefficient - 2.5377
- **Last Notable Activity_SMS Sent:** Coefficient - 2.0579

These attributes represent the most influential drivers of lead conversion. Optimizing these factors could significantly increase the conversion rate.

Thank You