



X Education - Lead Scoring Case Study

Identification of Hot Leads to focus more on them and thus enhancing the conversion ratio for X Education

Group members:

- 1. Sanket Badadal**
- 2. Sashi Kumar**
- 3. Sathwik Bidayath**

Background

X Education Company

- X Education , An education company named sells online courses to industry professionals
 - Many interested professionals land on their website
 - The company markets its courses on several websites 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
-

Background

X Education Company

- When these people fill up a form providing their email address or phone number, they are classified to be a lead
 - 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%
-

Problem Statement

X Education Company's
Problem

- X Education gets a lot of leads but its lead conversion rate is very poor
- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Problem Statement

X Education Company's
Problem

- We will help them to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- We are required to build a model wherein we need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be 80%.

Lead – Conversion Process

Lead to
Conversion
process

Lead Generation:

1. Ads on websites like Google
2. Referrals

Visit to X
Education
website by these
potential
customers
(professionals)

Visitors either
provide Email id
& Contact Details
Or
View videos etc

Tele calling and
Emailing activity
to all the leads

~30% leads get
converted

Proposed Solution:
A model to filter leads
so that leads to
conversion ratio is
80%+

Proposed Solution

Selection of Hot Leads

Leads Clustering

We cluster the leads into certain categories based on their tendency or probability to convert, thus, getting a smaller section of hot leads to focus more on.

Communicating with Hot Leads

Focus Communication

Since we would have a smaller set of leads to have communication with, we might make more impact with effective communication.

Conversion of Hot Leads

Increase conversion

Since we focussed on hot leads, which were more probable to convert, we would have a better conversion rate, and hence we can achieve the 80% target.

Solution

Selection of Hot Leads

For our Problem Solution, the crucial part is to accurately identify hot leads.

The more accurate we obtain the hot lead, the more chance we get of higher conversion ratio.

Since we have a target of 80% conversion rate, we would want to obtain a high accuracy in obtaining hot leads.

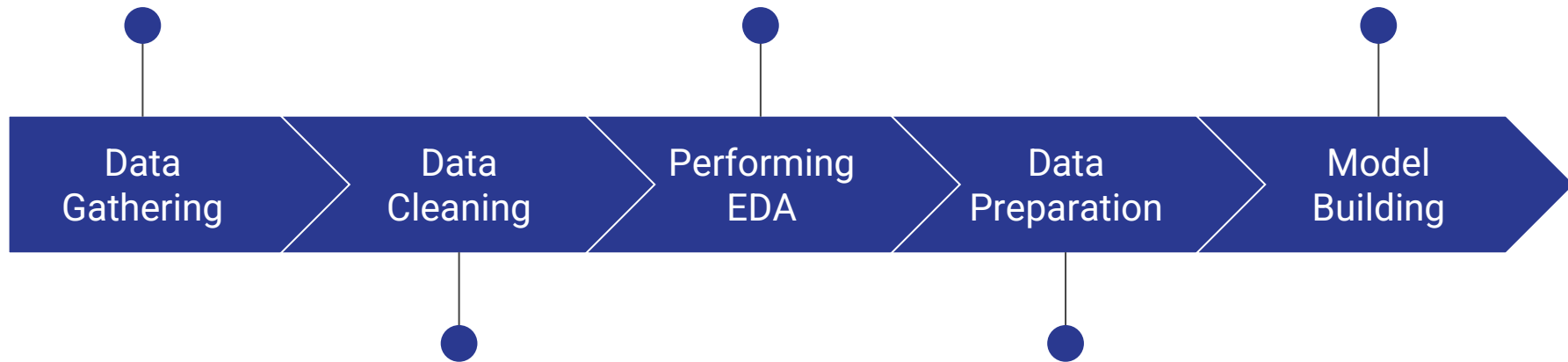
Implementation



Loading & Observing
the past data provided
by the Company

Univariate, Bivariate, and
Heatmap for numerical
and categorical columns

Performing pre-
requisites for RFE and
Logistic Regression



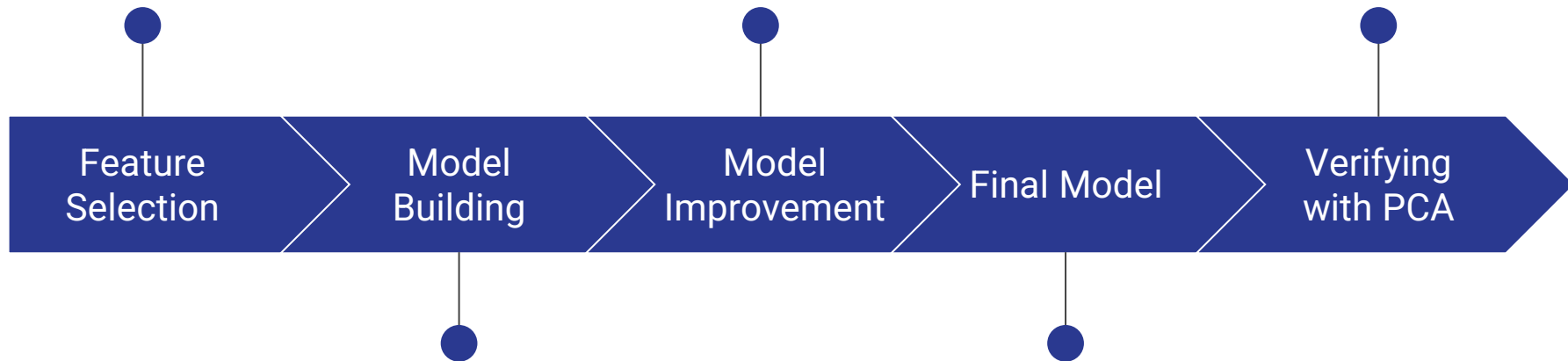
Duplicate removal, null value
treatment, unnecessary
column elimination, etc.

Outlier Treatment,
Feature-
Standardization

Selection of top 25
features using RFE

Reduction of columns
and Model re-building

Verifying our Final Model
Accuracy etc. with model
built with PCA

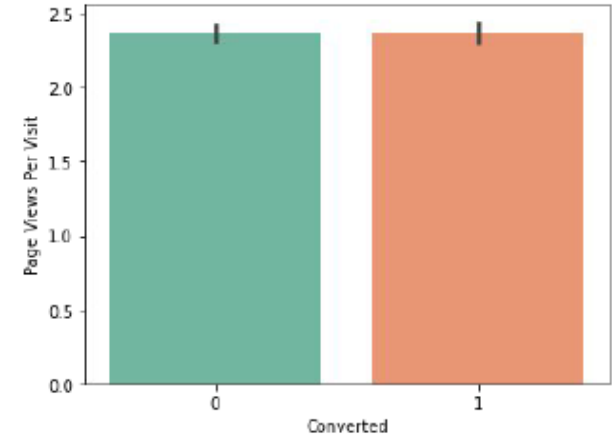
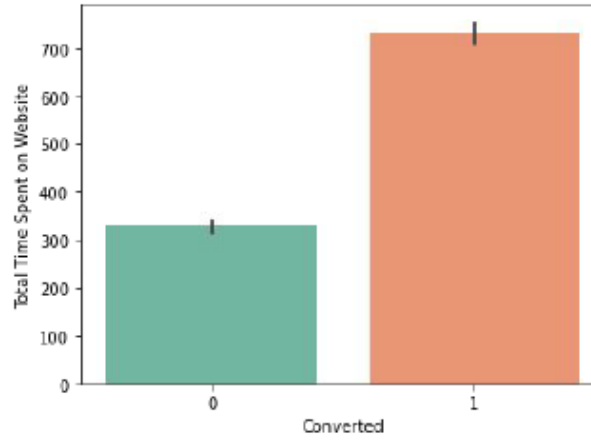
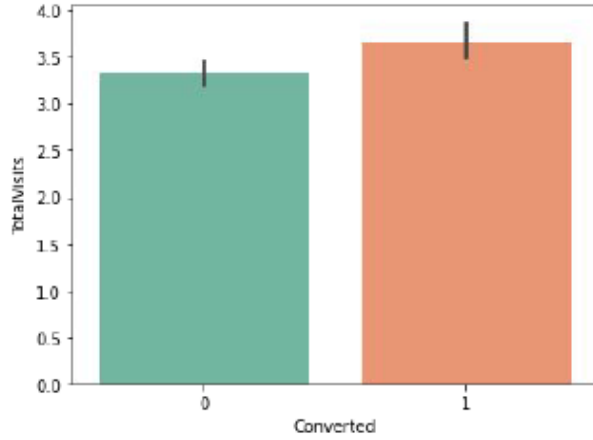


Model building using RFE for
selected columns

Final Model Analysis and
performance on Test
Data

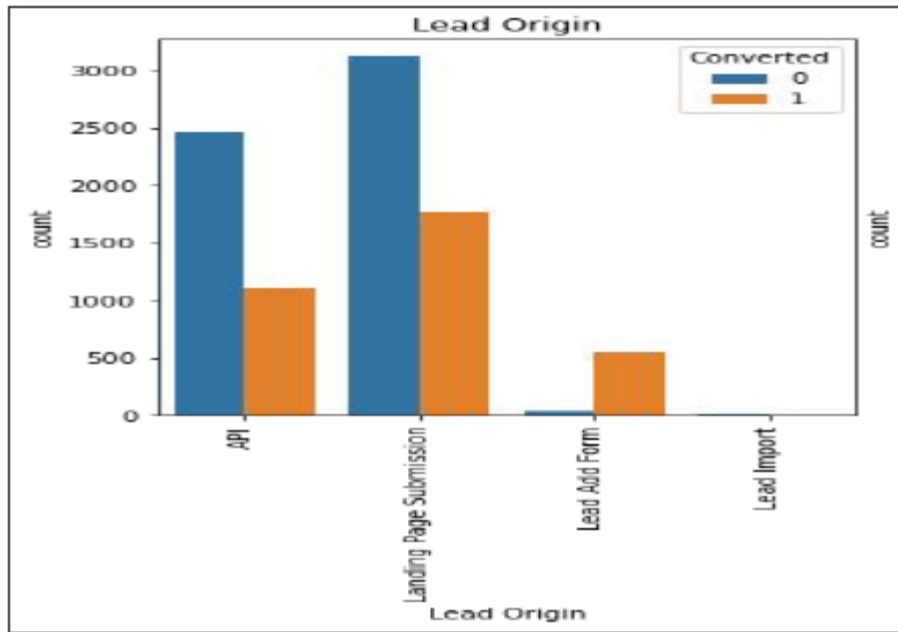


Plots (Visualization)

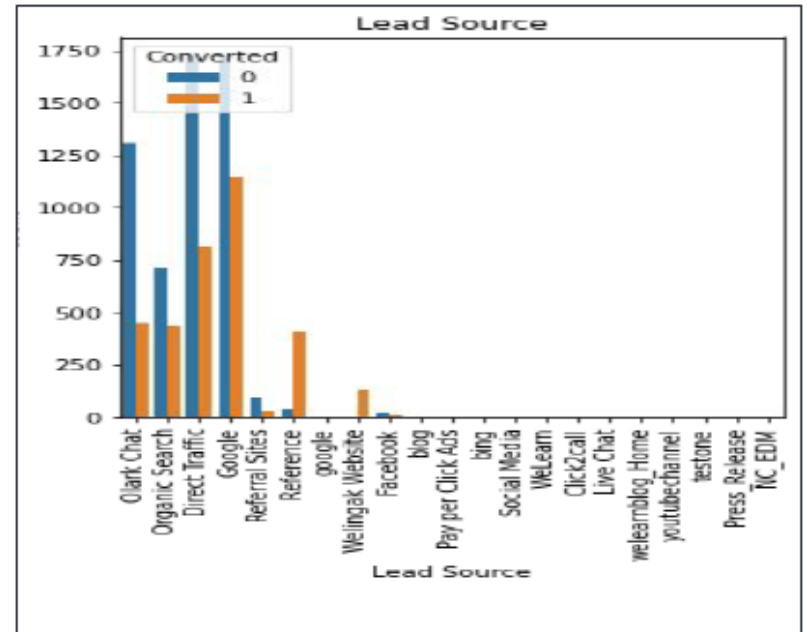


- The conversion rates were high for Total Visits, Total Time Spent on Website and PageViews Per Visit

Analysis on Total Visits, Total time & Page views per visit

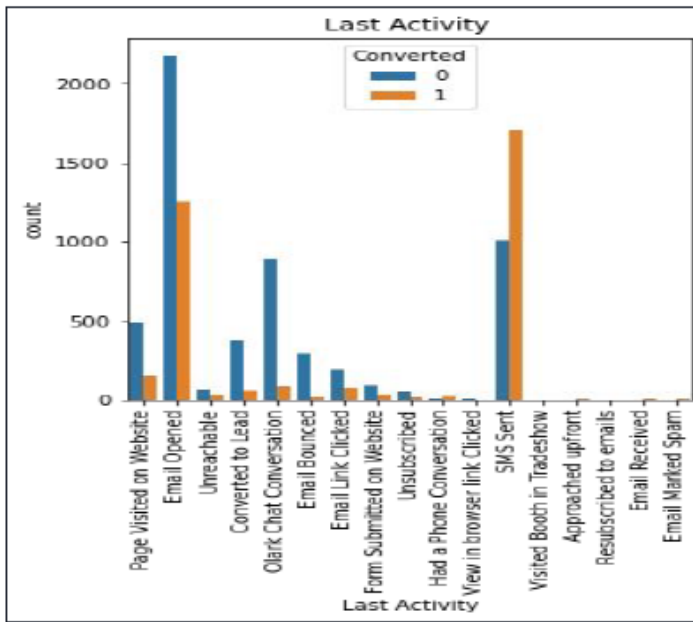


- In Lead Origin, maximum conversion happened from Landing Page Submission
- But if we saw percentage wise than lead Add Form is highest

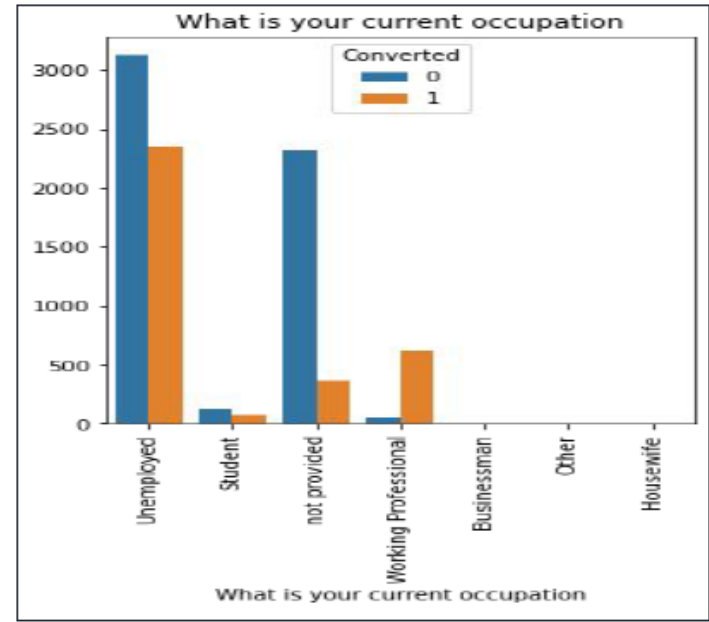


- We can see that maximum lead comes from Google & Direct Traffic
- Reference leads are most important.

Analysis on Lead origin & Lead source



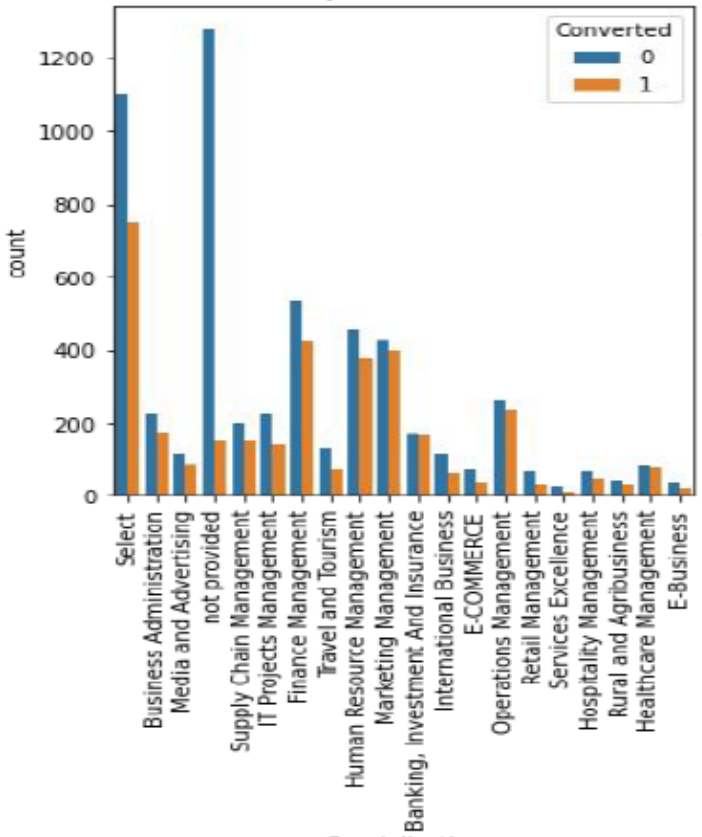
- Where last activity SMS send have highest percentage of conversion.
- Last Activity Email Opened is the 2nd highest of conversion.



- Working Professional conversion rate is very high.
- Maximum Unemployed peoples are converted.

Analysis on Last Activity & current occupation

Specialization

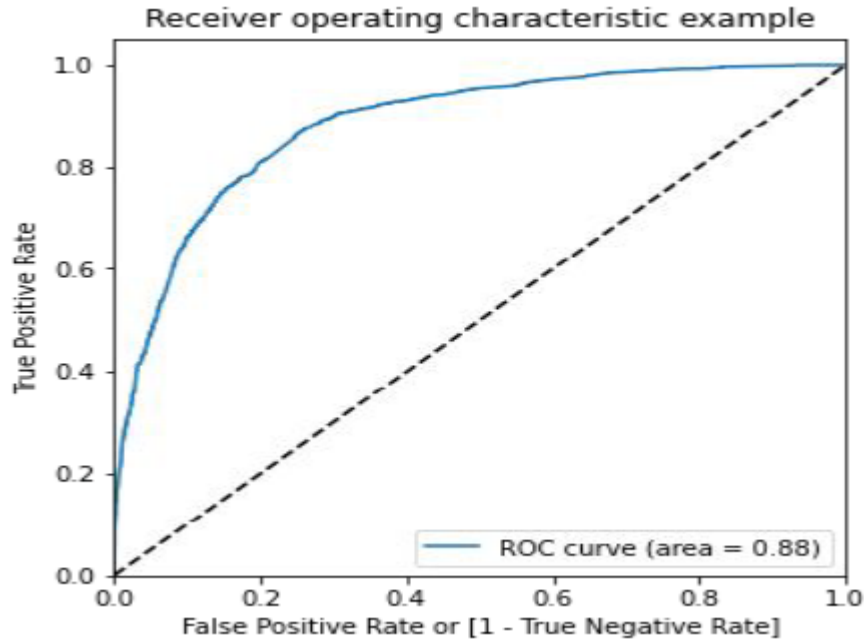


- We must stop calling to those customer for whom Specialization “Not Provided” because conversion rate is very low.
- Specialization with highest conversion:-
 - 1) Marketing Management
 - 2) Human Resource Management
 - 3) Finance Management
 - 4) Operations Management
 - 5) Business Administration
 - 6) Banking, Investment & Insurance
- We must focus this specialization

Analysis on specialization

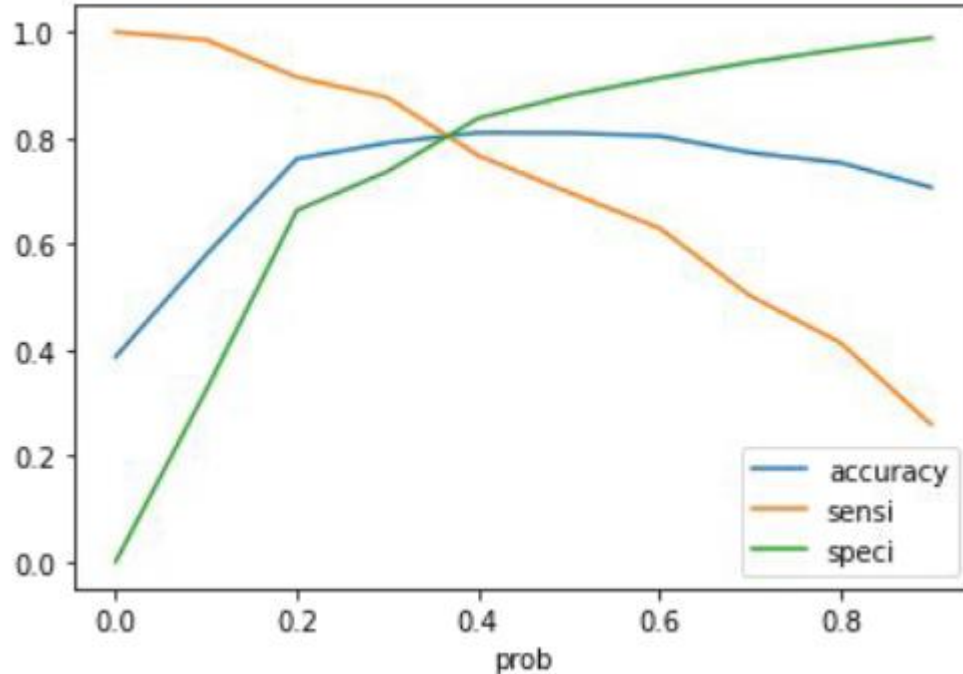
Variables Impacting the conversion Rate

1. Do Not Email
2. Total Visits
3. Total Time Spent On Website
4. Lead Origin – Lead Page Submission
5. Lead Origin – Lead Add Form
6. Lead Source - Olark Chat
7. Last Source – Welingak Website
8. Last Activity – Email Bounced
9. Last Activity – Not Sure
10. Last Activity – Olark Chat Conversation
11. Last Activity – SMS Sent
12. Current Occupation – No Information
13. Current Occupation – Working Professional
14. Last Notable Activity – Had a Phone Conversation
15. Last Notable Activity - Unreachable



- After building the final model making prediction on it(on trainset), we created ROC curve to find the model stability with AUC score(area under the curve)
- As we can see from the graph plotted on the right side, the area score is 0.88 which is a great score.
- And our graph is leaned towards the left side of the border which means we have good accuracy.

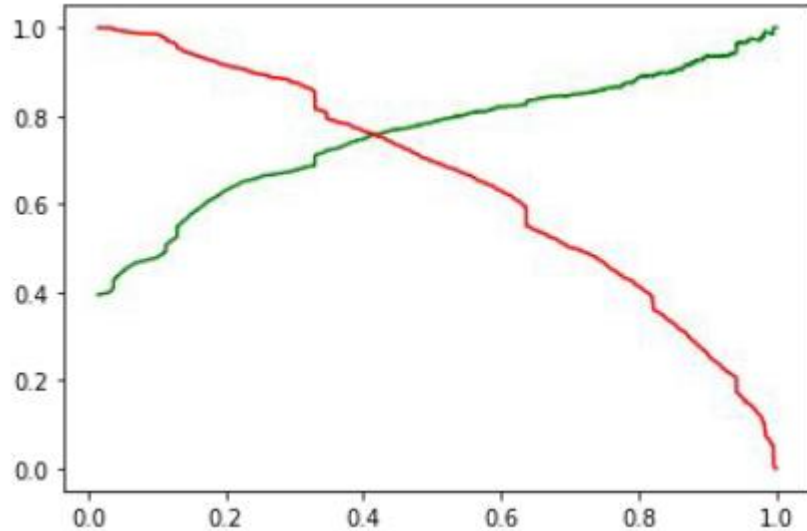
Evaluating the model



- We found that on 0.38 Cut-off point all the score of accuracy, sensitivity and specificity are in a close range which is the ideal point to select and hence it was selected.

- Accuracy - 80 %
- Sensitivity - 79 %
- Specificity - 80 %

Model Evaluation – Sensitivity and Specificity on Train data set



The graph depicts an optimal cut off
0.41

based on Precision and Recall.

→ Precision - 75 %

→ Recall - 76 %

Model Evaluation – Precision and Recall on Train Dataset

Confusion Matrix

1472	272
228	751

- Accuracy - 81 %
- Specificity- 84 %
- Precision - 73 %
- Recall - 76 %

Model Evaluation – Accuracy, Precision and Recall Test Dataset

Conclusion:

- Accuracy, Sensitivity and Specificity values of test set are around 81%, 84% and 76% which are approximately closer to the respective values calculated using trained set.
- The Accuracy, Precision and Recall score we got from test set are in acceptable range.
- This model has an ability to adjust with the company's requirements in coming future.
- the variables that mattered the most in the potential buyers are as per below -
 - 1) The total time spent on the Website.
 - 2) Total Visits
 - 3) Last activity was:- a. SMS, b. Olark chat conversation
 - 4) What is your current occupation Working Professional
 - 5) Lead Origin Lead Add Form