



LEAD SCORING CASE STUDY

Naveen Babu S

Naveen Kumar GT

Neha Shailendra




PROBLEM STATEMENT

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some video, they are classified to be a lead.
- Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.
- As you can see, there are a lot of leads generated in the initial stage but only a few of them come out as paying customers.
- X Education has appointed you to help them select the most promising leads. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.



BUSINESS GOAL

- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.
 - A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.
- 

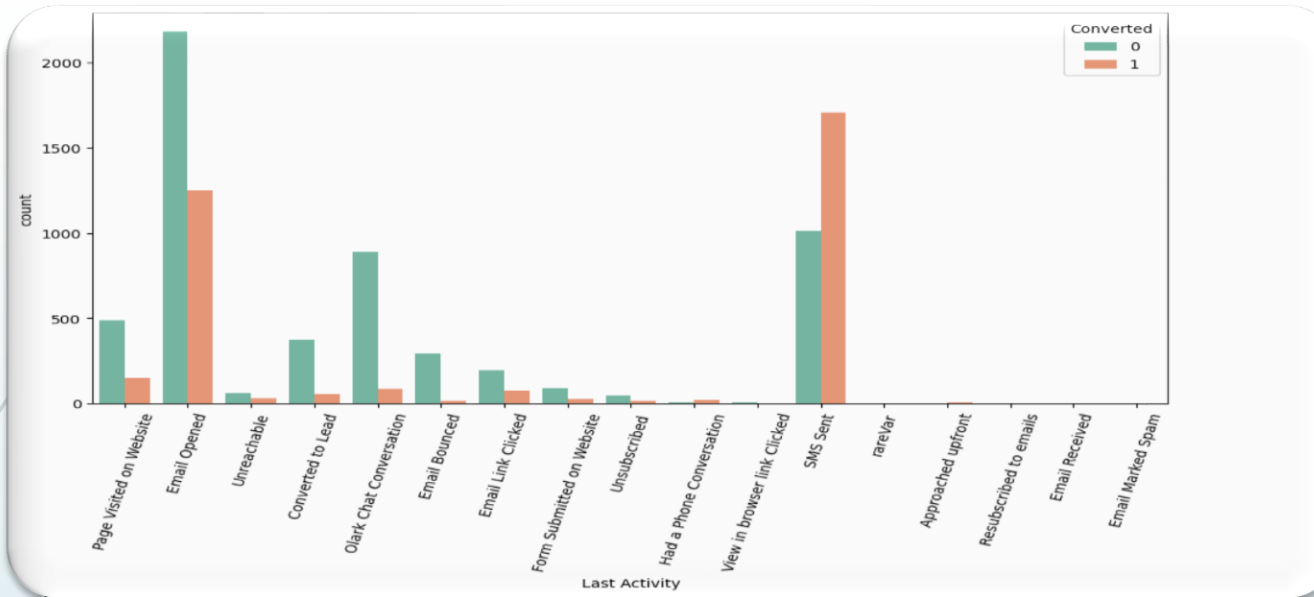


STEPS INVOLVED

- Business Problem Understanding
- Data Sourcing
- Data Understanding and cleaning
- Exploratory Data Analysis
- Feature Scaling
- Data Preparation for Model Building
- Build the Logistic Regression Model and predict target variable
- Assign the Lead Score for each leads
- Evaluation the model by different metrics
- Finalize the best model for prediction



EXPLORATORY DATA ANALYSIS

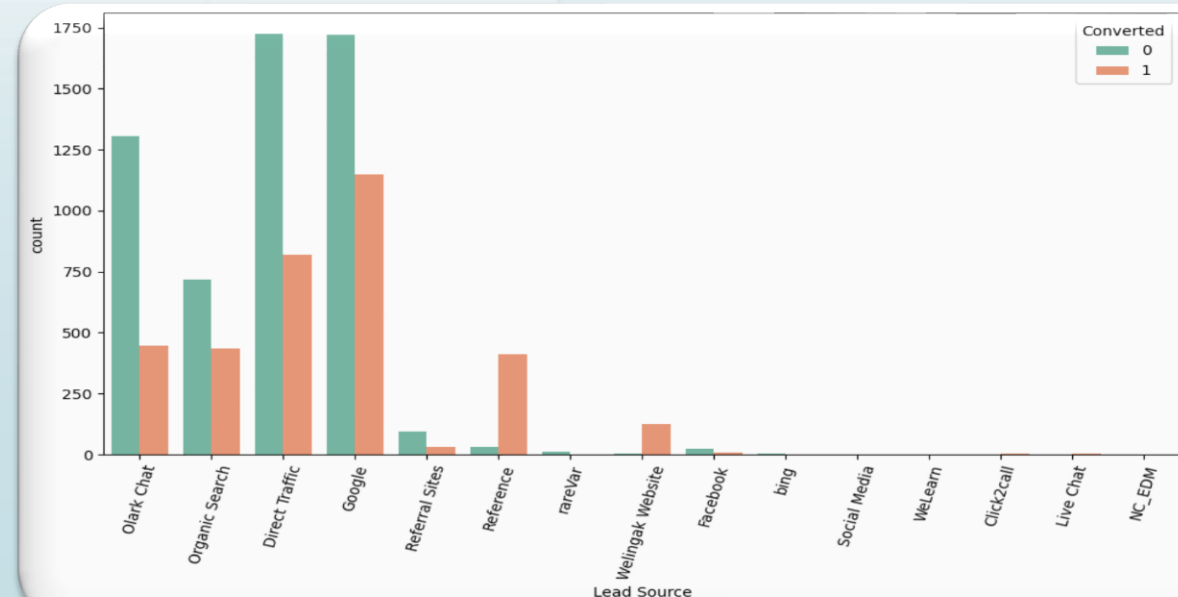


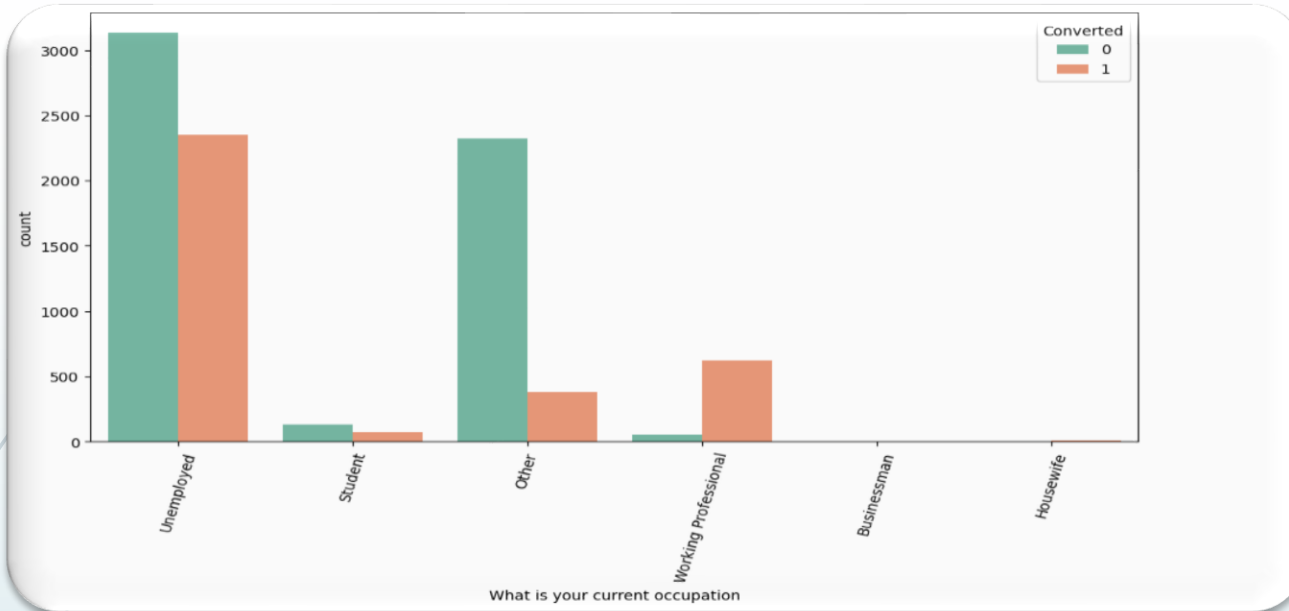
Last Activity

- ✓ Maximum Leads are generated having the last activity as E-mail opened but the conversion rate is not so good.
- ✓ SMS send as the last activity has high conversion rate.

Lead Source

- ✓ Maximum Leads are generated by Google and Direct Traffic
- ✓ Conversion rate of Reference leads and Welinkgak Website leads is very high.



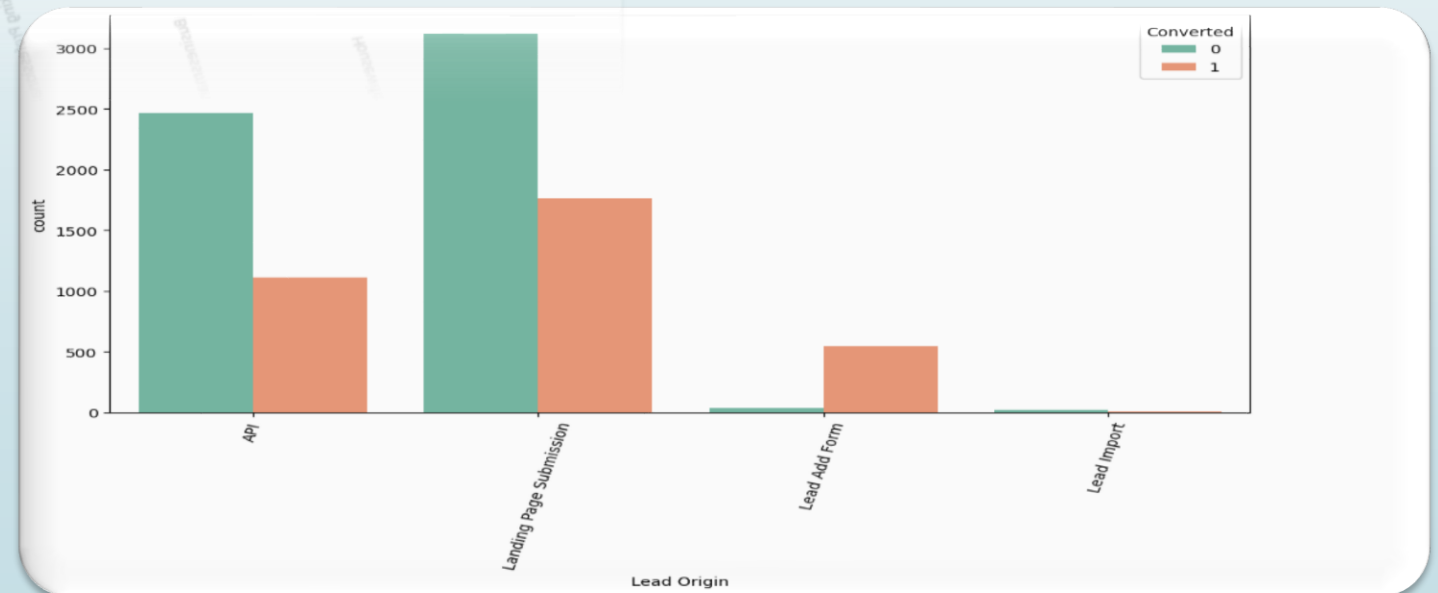


Occupation

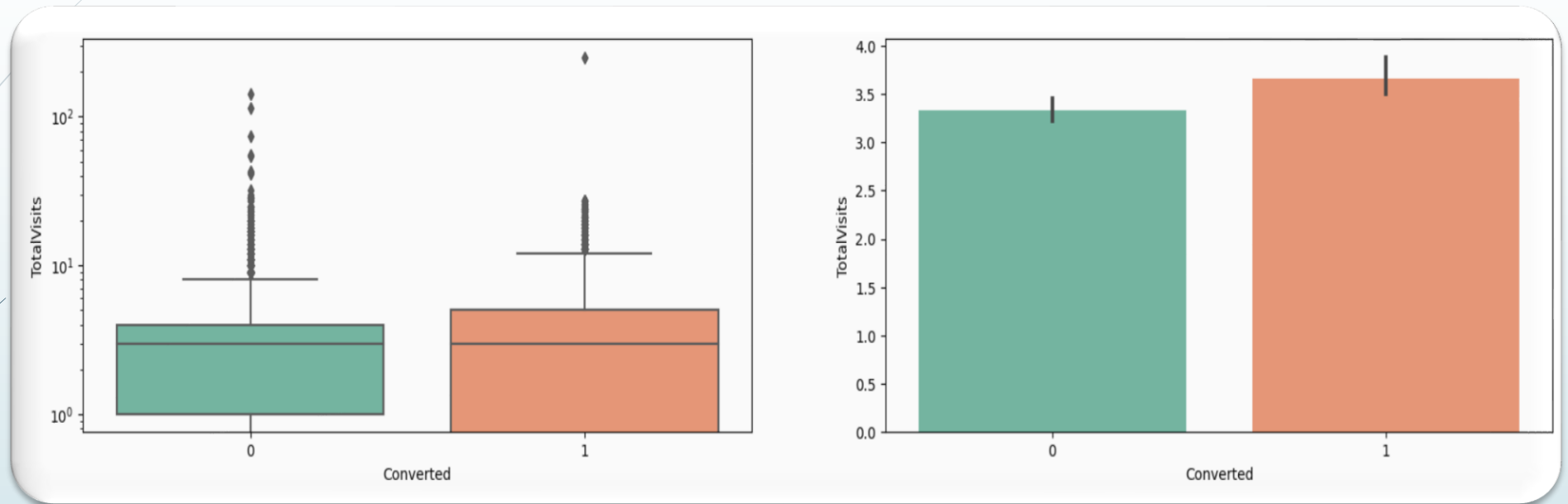
- ✓ Maximum leads generated are Unemployed and their conversion rate is more than 70%.
- ✓ Conversion rate of working professionals is very high.

Lead Origin

- ✓ Origin of API and landing page submission makes the most of leads but conversion rate is low.

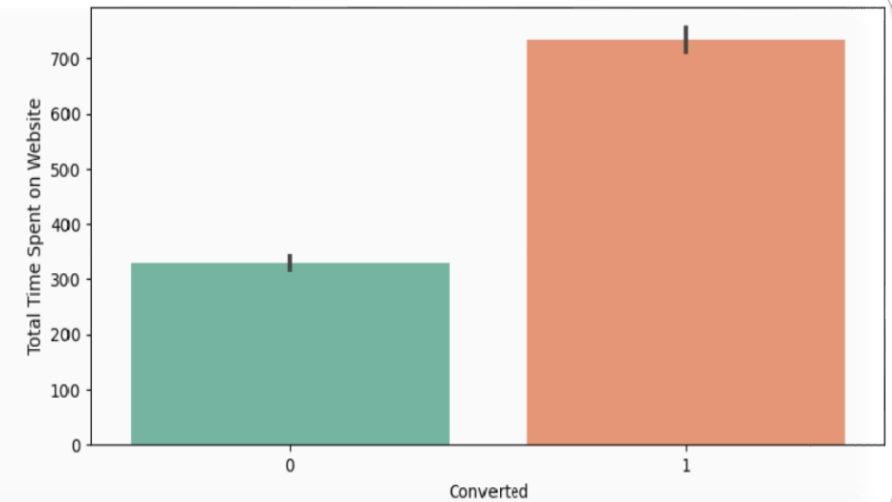
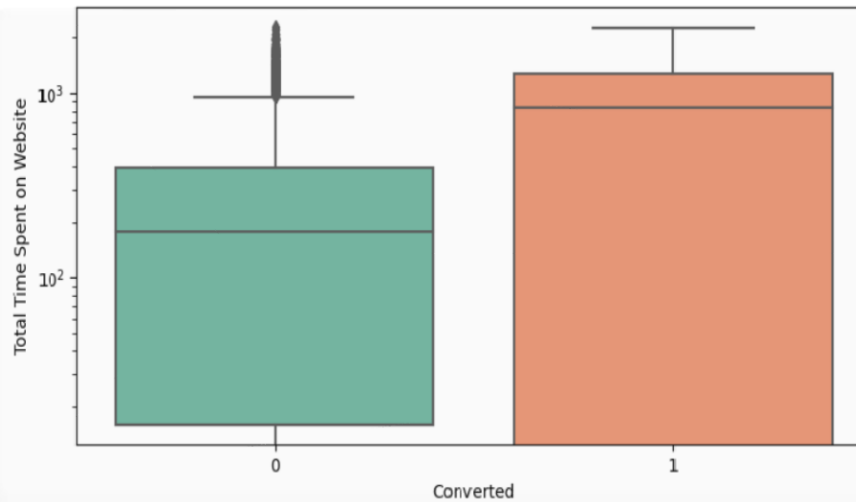


Total Visits



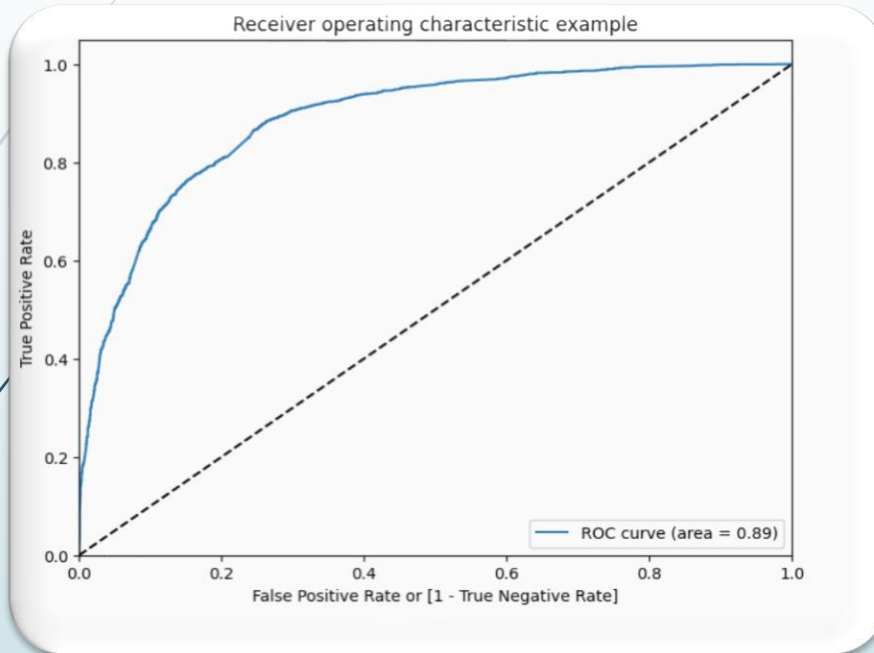
- ✓ The median for both converted and not converted leads are same.
- ✓ Higher Total visits have slightly higher chance of being promising lead.

Total Time Spend on Website

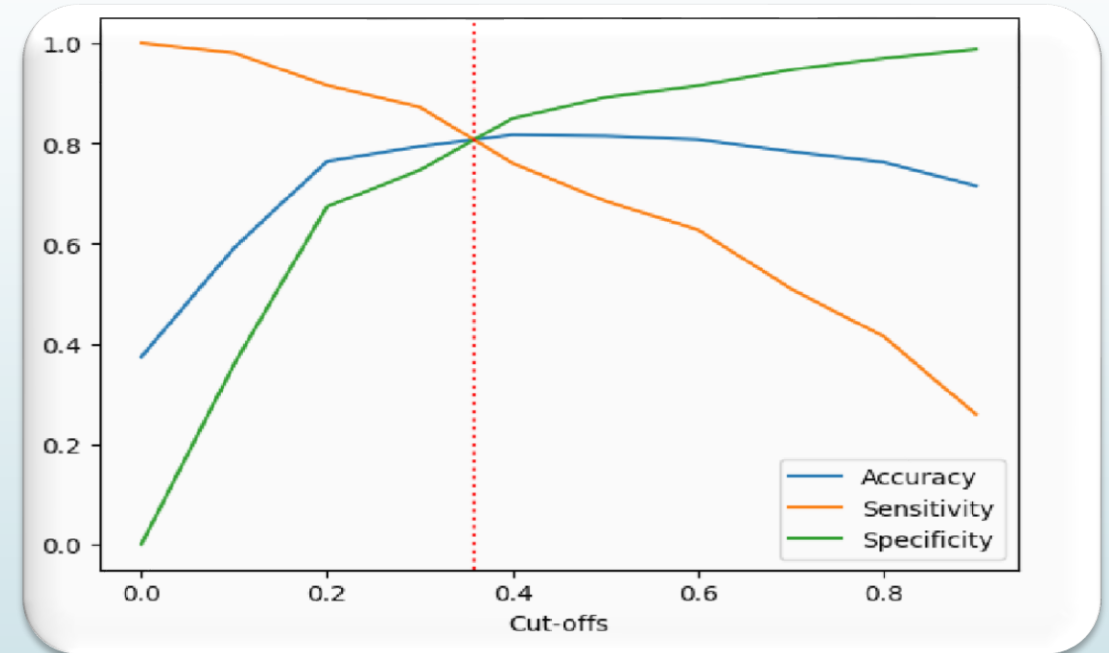


- ✓ As we can see, leads spending more time on website are more likely to convert, thus website should be made more engaging to increase conversion rate.

MODEL BUILDING AND EVALUATION

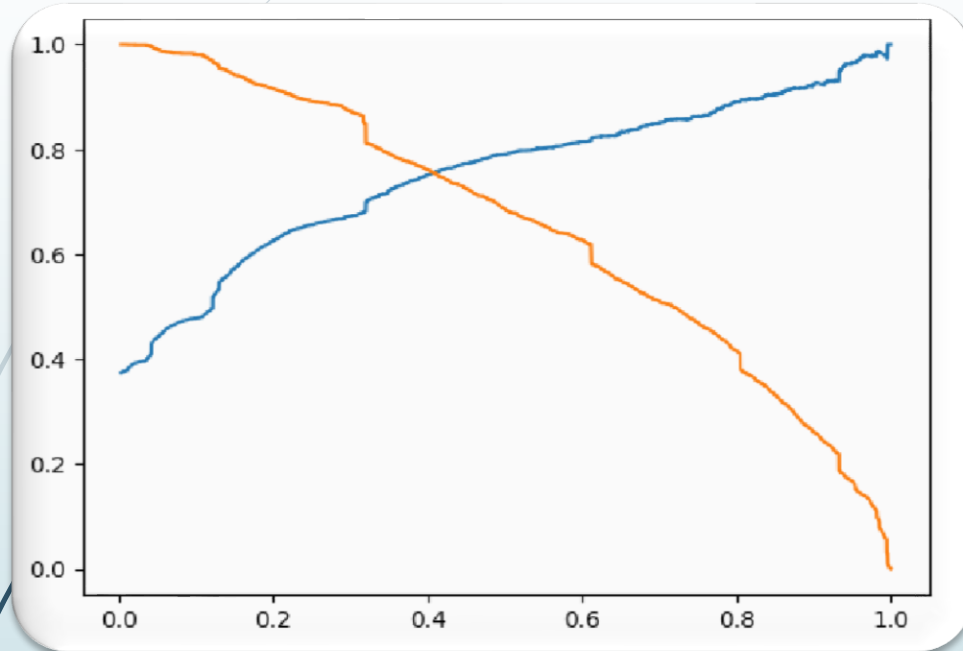


Area Under Curve \rightarrow 0.89



Optimal cut-off \rightarrow 0.358

Evaluation Metrics – Train



Confusion matrix:

3275

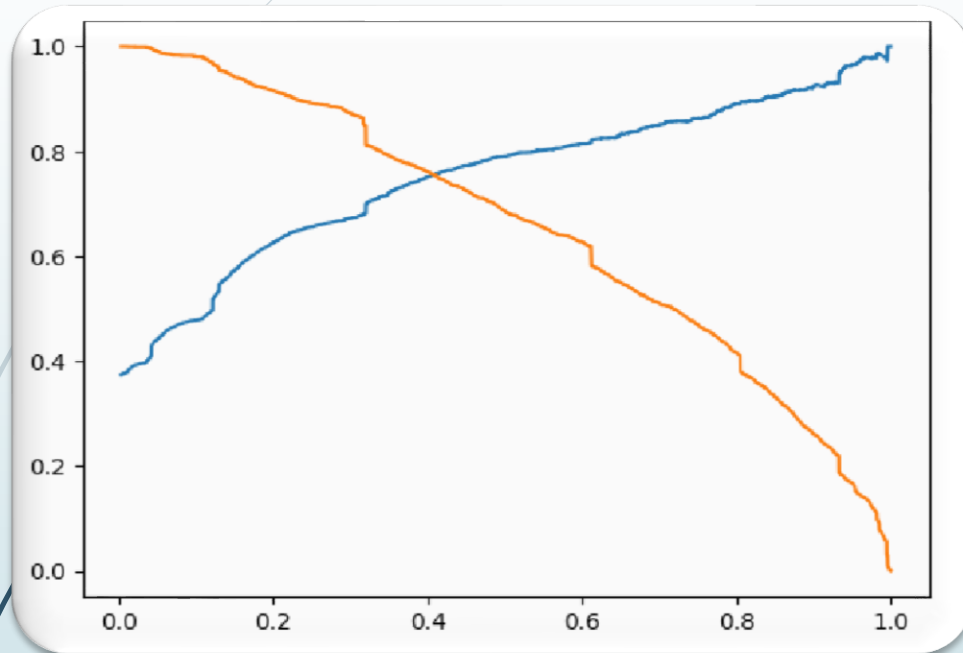
702

511

1863

- Accuracy --> 80.9 %
- F-1 Score --> 75.44 %
- Precision --> 72.63 %
- Sensitivity --> 78.48 %
- Specificity --> 82.35 %

Evaluation Metrics – Test



Confusion matrix:

1357	305
236	825

- Accuracy --> 80.13 %
- F-1 Score --> 75.31 %
- Precision --> 73.01 %
- Sensitivity --> 77.76 %
- Specificity --> 81.65 %

Evaluation Metrics - Overall

Confusion matrix:

4632	1007
747	2688

- Accuracy --> 80.67 %
- Sensitivity --> 78.25 %
- Specificity --> 82.14 %

HOT LEADS: Out of 9074 leads only 1601 are most potential leads.

FINAL FEATURES OF OUR MODEL

Lead Source_Welingak Website	6.267811
What is your current occupation_Working Professional	3.790313
Lead Source_Reference	3.522839
Last Notable Activity_Had a Phone Conversation	3.239110
Last Notable Activity_Unreachable	2.144772
What is your current occupation_Student	1.355173
Lead Source_Olark Chat	1.350775
What is your current occupation_Unemployed	1.224050
Last Activity_SMS Sent	1.209129
Total Time Spent on Website	1.126691
Last Activity_Olark Chat Conversation	-1.145573
Last Activity_Email Bounced	-1.200291
Last Activity_Converted to Lead	-1.247763
Do Not Email	-1.291296



RECOMMENDATION

The Company should target the leads from

■ **Lead Source**

- Welingak Website
- Reference

■ **Occupation**

- Working Professional
- Unemployed

■ **Last Activity**

- Had the phone Conversation
- SMS send

THANKYOU