Lead Scoring Case Study

Submitted by

Kuppala Sri Sai Raviteja

Neel Panchal

Contents

- Problem Statement
- Problem Approach
- EDA
- Model Evaluation
- Observations
- Conclusion

Problem Statement

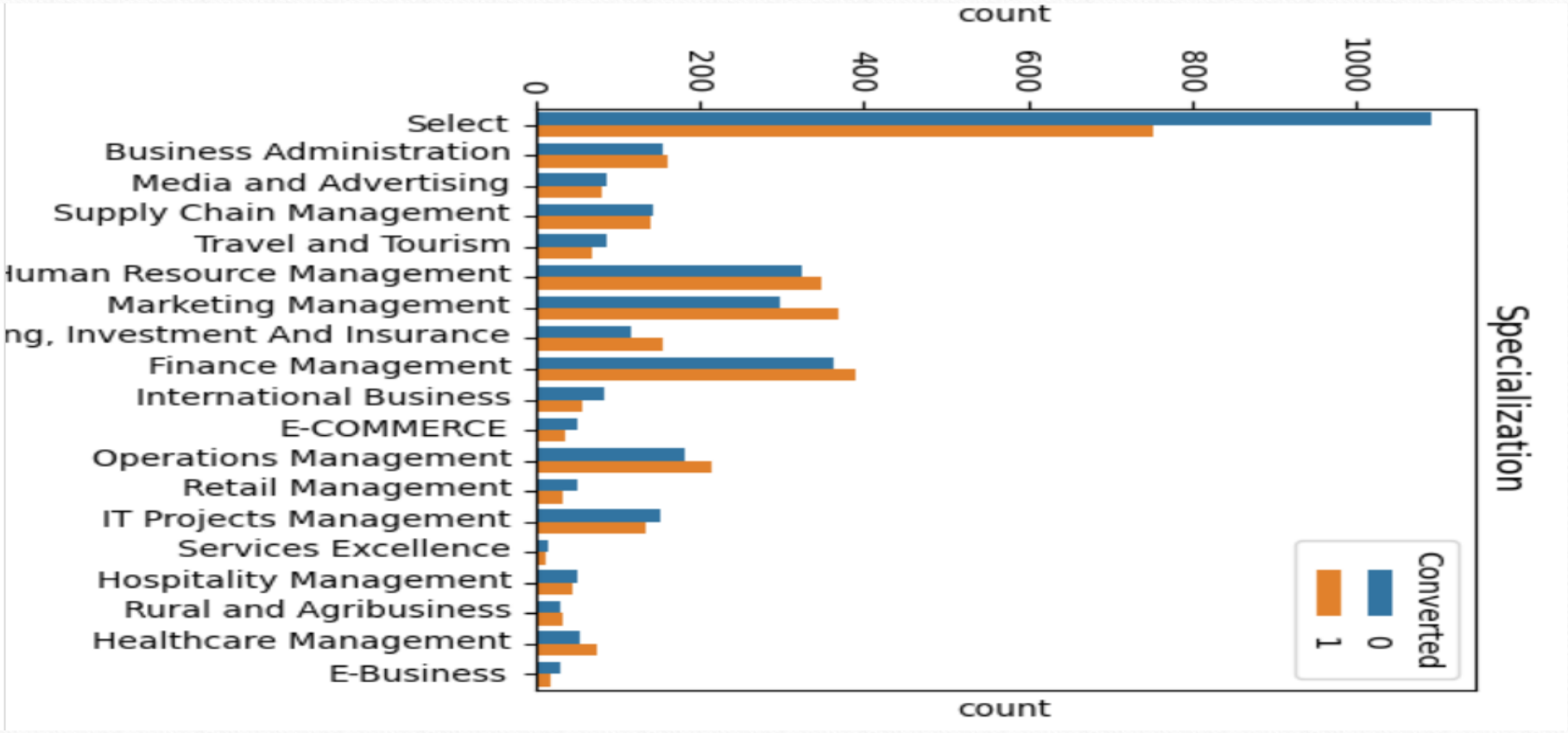
- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the 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.

Business Objective

- Build a 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.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.
- Adjust the deployment of the model for the future use.

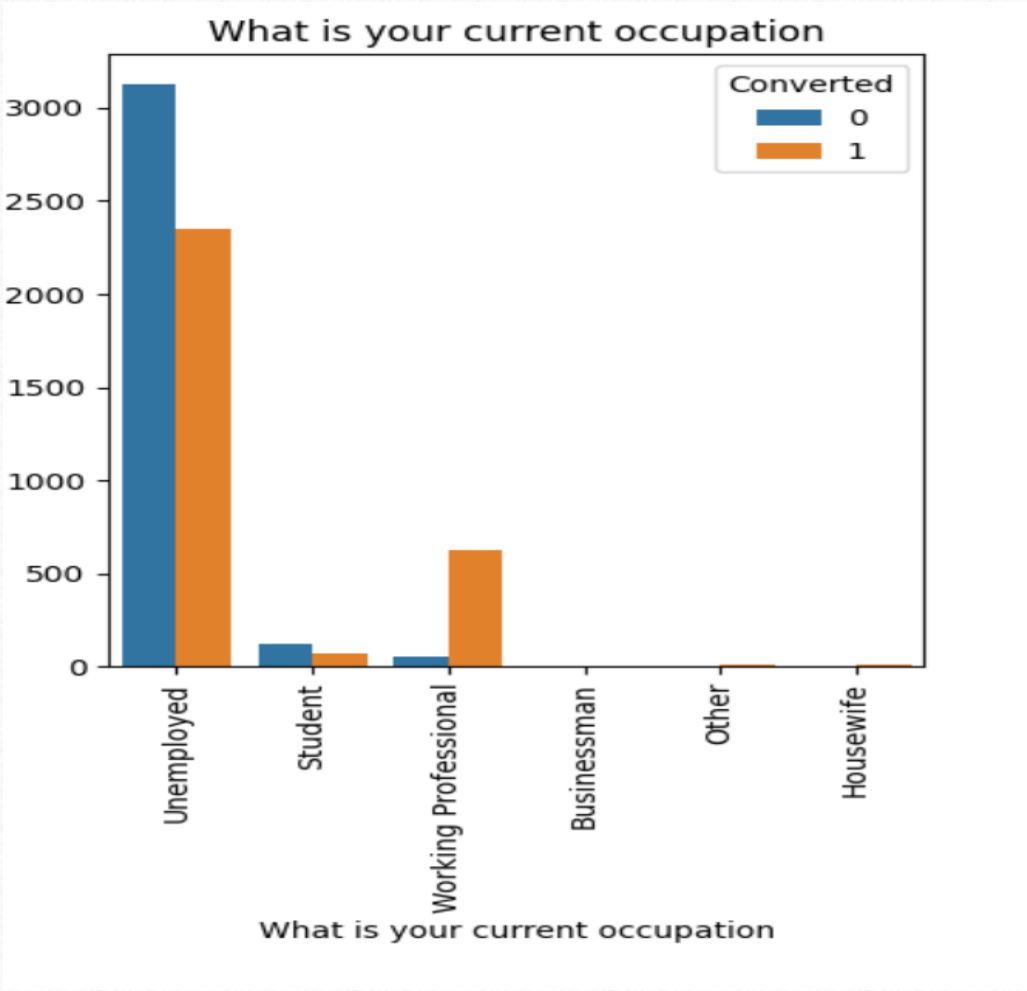
Problem Approach

- Data Sourcing, Cleaning and Preparation
- Feature Scaling
- Train-Test-split
- Model Building(RFE Rsquared VIF and p-values)
- Model Evaluation
- Making Predictions on Test set

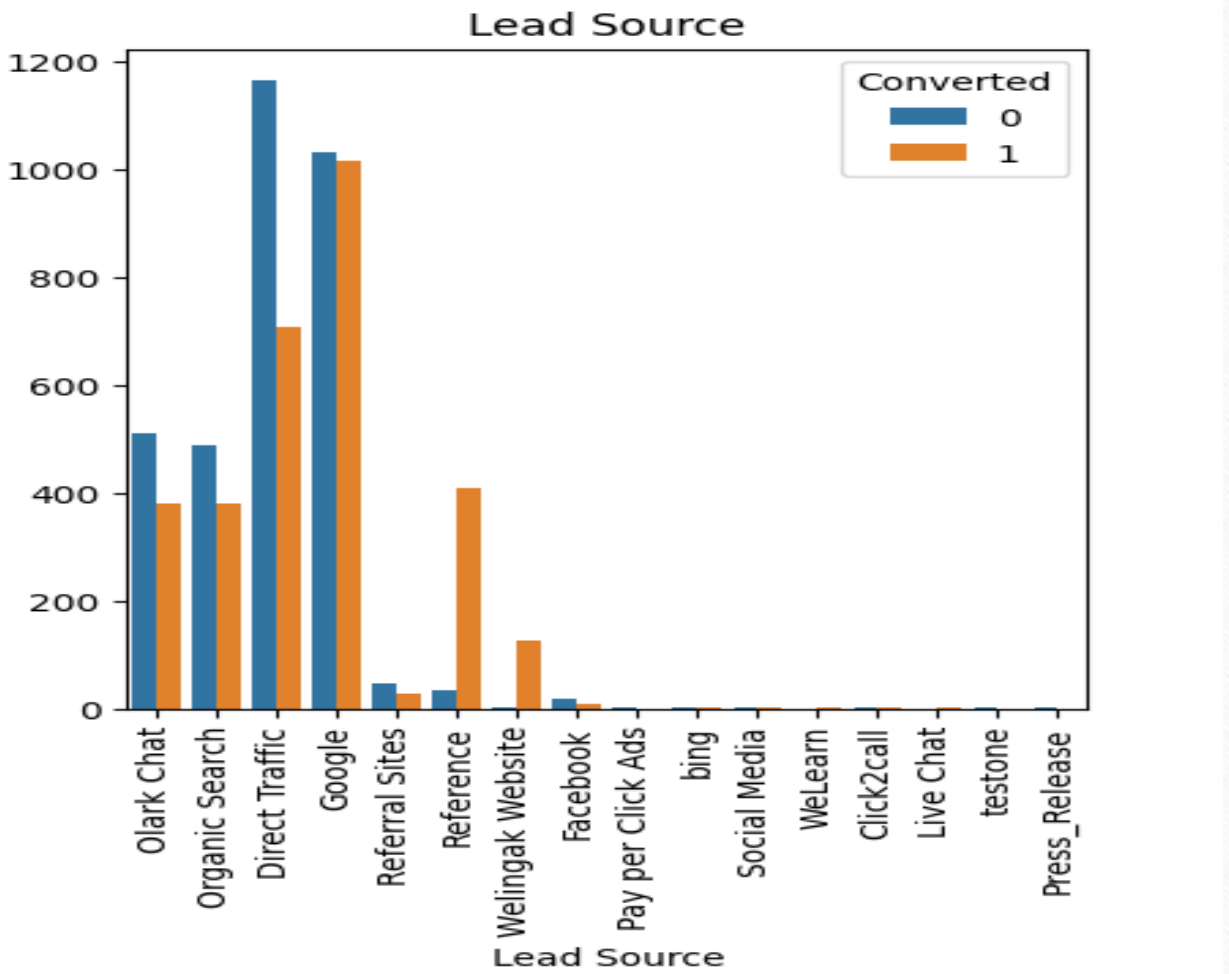


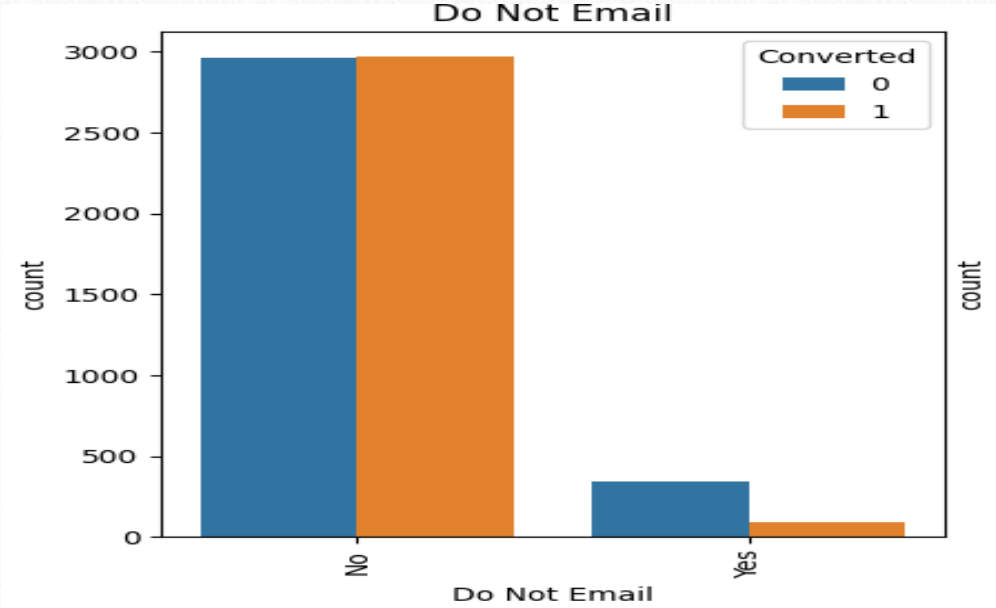
We see that specialization with Management in them have higher number of leads as well as leads converted. So, this is definitely a significant variable and should not be dropped.

Working Professionals going for the course have high chances of joining it. Unemployed leads are the most in terms of Absolute numbers.

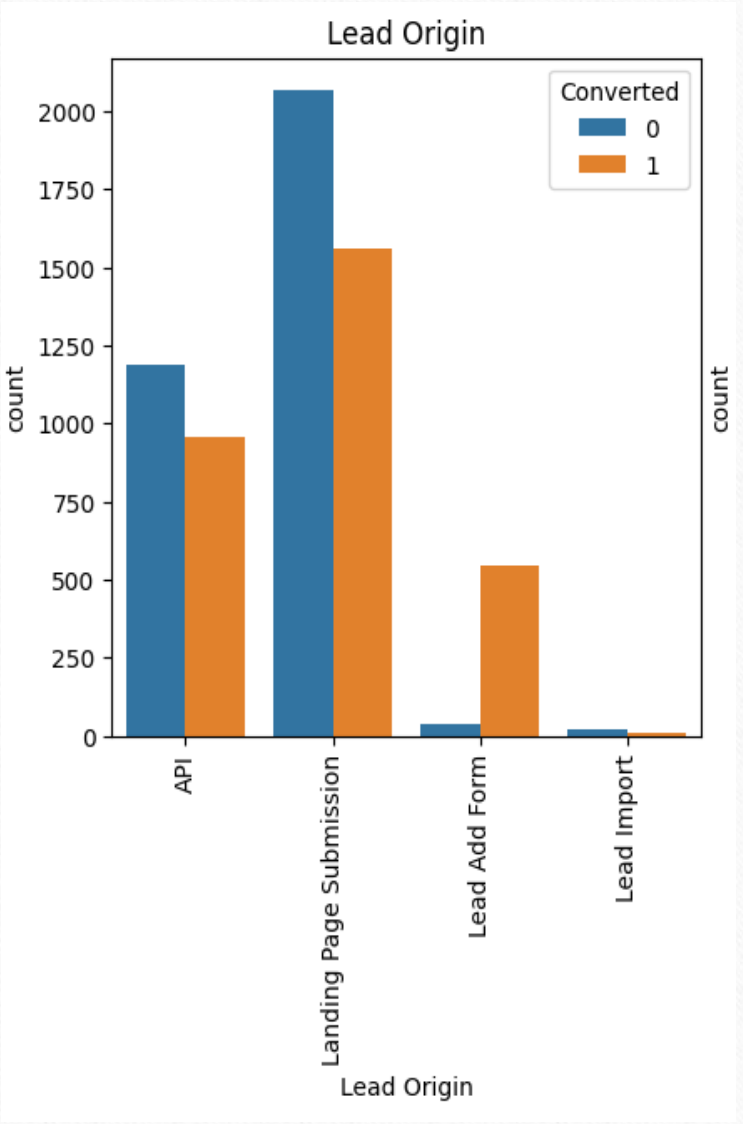


1. Maximum number of leads are generated by Google and Direct traffic.
2. Conversion Rate of reference leads and leads through welingak website is high.
3. To improve overall lead conversion rate, focus should be on improving lead conversion of Olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.





1. API and Landing Page Submission bring higher number of leads as well as conversion.
2. Lead Add Form has a very high conversion rate but count of leads are not very high.
3. Lead Import and Quick Add Form get very few leads.
4. In order to improve overall lead conversion rate, we have to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.



Model Evaluation – Sensitivity and Specificity on Train Data Set

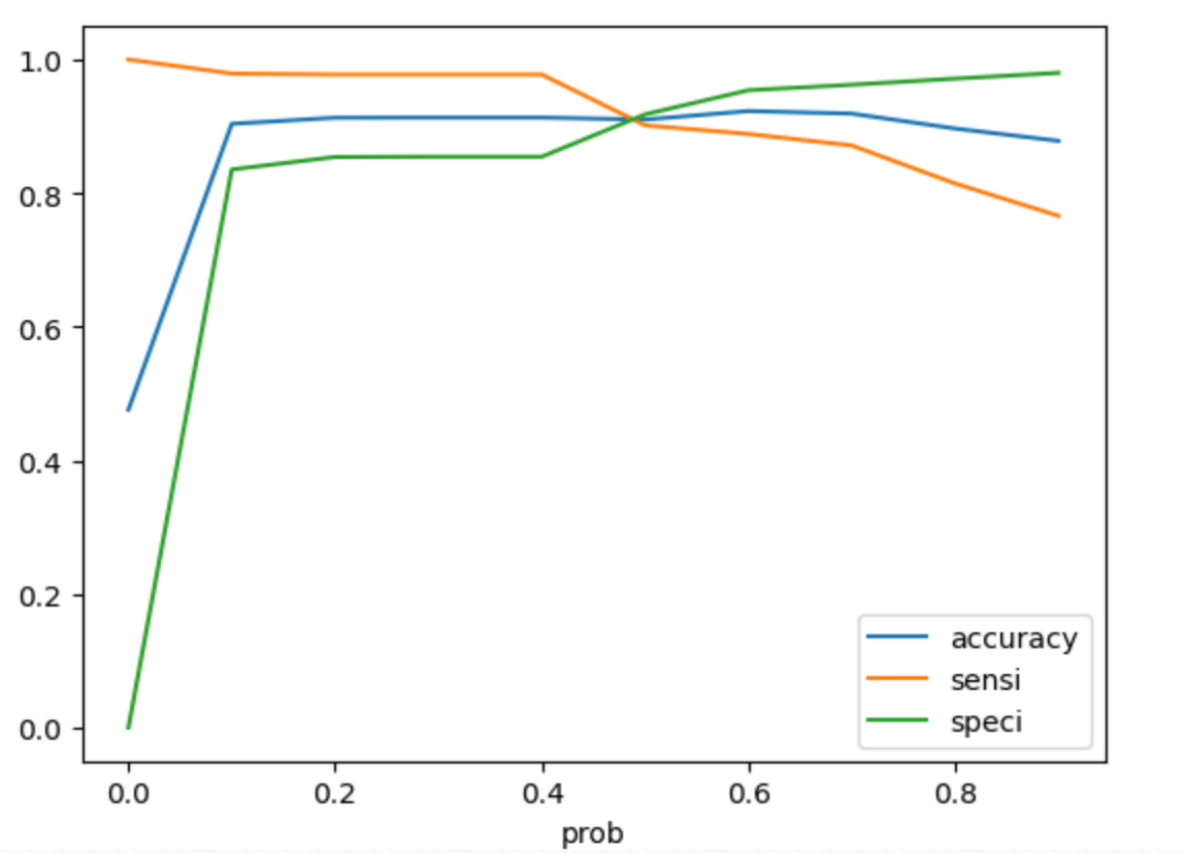
Observation:

So, as we can see above the model seems to be performing well. The ROC curve has a value of 0.97, which is very good. We have the following values for the Train Data:

Accuracy : 90.73%

Sensitivity : 91.00%

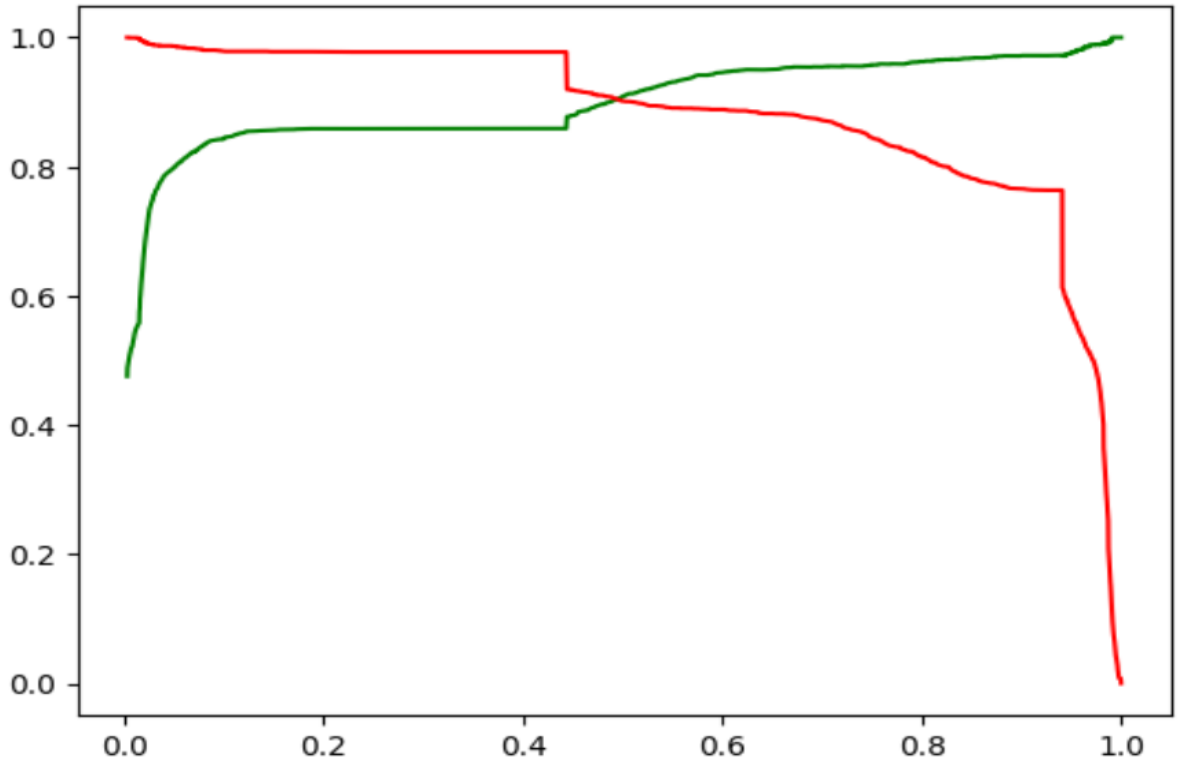
Specificity : 90.48%



As you can see that around 0.48, you get the optimal values of the three metrics. So ,let's choose 0.48 as our cutoff now.

Model Evaluation – Precision and Recall on Train Data Set

Precision : 89.67%
Recall: 91.00%



The graph depicts an optimal cutoff 0.48 based on precision and Recall.

Model Evaluation – Sensitivity and Specificity on Test Data Set

Observation:

After running the model on the test data these are the figures we obtain

Test Data:

Accuracy : 90.33%

Sensitivity : 91.55%

Specificity : 89.15%

Conclusion

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.

- Accuracy, Sensitivity and Specificity values of test set are around 90%, 91% and 90.48% which are approximately closer to the respective values calculated using trained set.

- The top variables that contribute for lead getting converted in the model are

1. Total time spent on website
2. Lead Source_Welingak Website
3. Tags_Will revert after reading the email

- Hence overall this model seems to be good.