

# Lead Scoring Case Study

---

Yash Kharte

# BUSINESS 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.

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%.

# BUSINESS OBJECTIVE

The company requires us to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead scores have a higher conversion chance and the customers with lower lead scores have a lower conversion chance.

The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

# MODEL BUILDING

We can easily find insignificant columns in our csv.

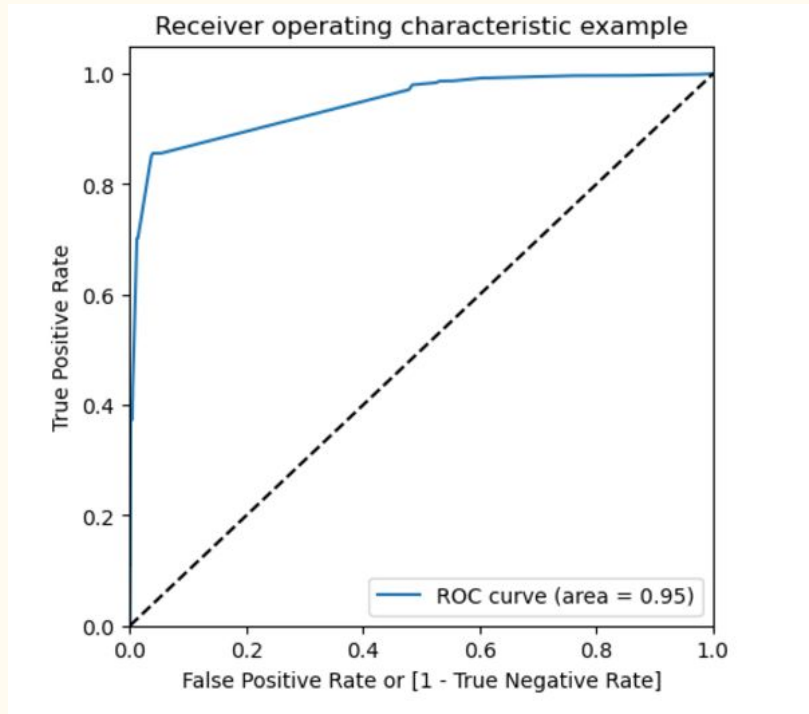
<b>Dep. Variable:</b>	Converted	<b>No. Observations:</b>	6351
<b>Model:</b>	GLM	<b>Df Residuals:</b>	6265
<b>Model Family:</b>	Binomial	<b>Df Model:</b>	85
<b>Link Function:</b>	Logit	<b>Scale:</b>	1.0000
<b>Method:</b>	IRLS	<b>Log-Likelihood:</b>	nan
<b>Date:</b>	Tue, 19 Nov 2024	<b>Deviance:</b>	71012.
<b>Time:</b>	22:47:39	<b>Pearson chi2:</b>	3.47e+18
<b>No. Iterations:</b>	100	<b>Pseudo R-squ. (CS):</b>	nan
<b>Covariance Type:</b>	nonrobust		

	coef	std err	z	P> z	[0.025	0.975]
const	1.325e+15	7.53e+07	1.76e+07	0.000	1.32e+15	1.32e+15
Do Not Email	-2.346e+14	4.45e+06	-5.27e+07	0.000	-2.35e+14	-2.35e+14

# EVALUATE MODEL

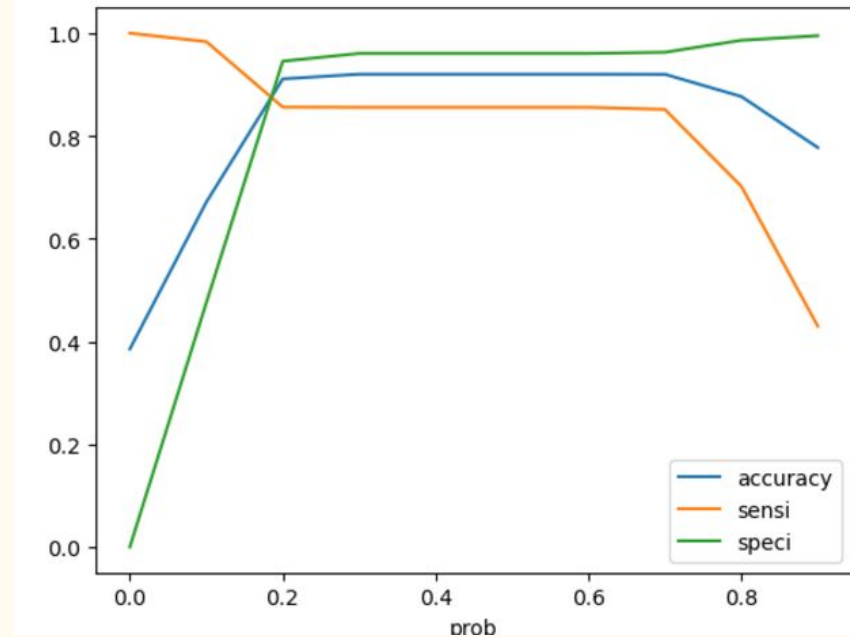
After constructing the final model and generating predictions on the training dataset, an ROC curve was created to assess the model's performance and stability. The area under the curve (AUC) score was calculated as 0.95, indicating a very strong model.

Moreover, the ROC curve leans towards the top-left corner, signifying high accuracy and effective predictive capability.



# FIND OPTIMAL CUT-OFF POINT

As seen in the figure,  
0.2 is the optimal cut  
off point



# CONCLUSION

In business terms, this model can adjust with the company's requirements very effectively.

Columns most relevant are :

- Total Time Spent on Website
- Lead Quality\_Not Sure
- Tags\_Will revert after reading the email

THANK YOU





