Lead Scoring Case Study

Kuldip Pathak ds49

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%.

Now, although 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. A typical lead conversion process can be represented using the following funnel:

Goals of the Case Study

- There are quite a few goals for this case study:

1. 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.

2. There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

STRATEGY

- IMPORT DATA

- Clean and prepare the data

- Exploratory data analysis

- Scaling

- Data model building

- Build a logistic regression model

- Test the model on train set the

- Measure accuracy of the model and other metrics for evaluation
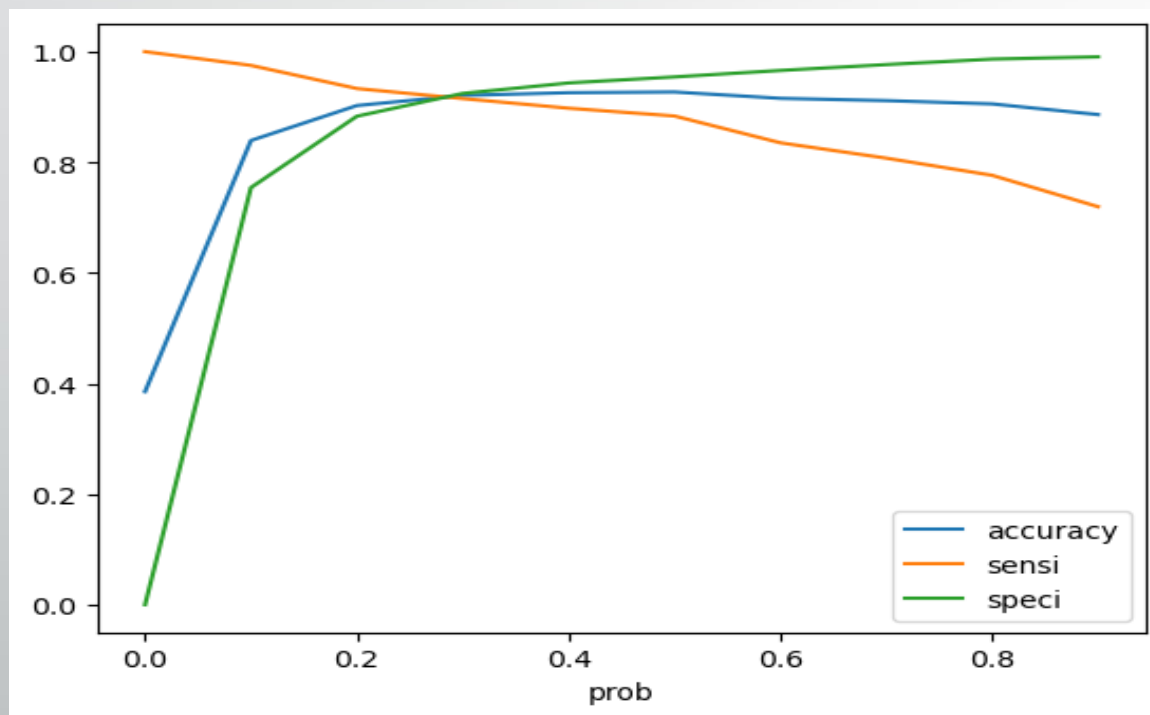
Exploratory data analysis
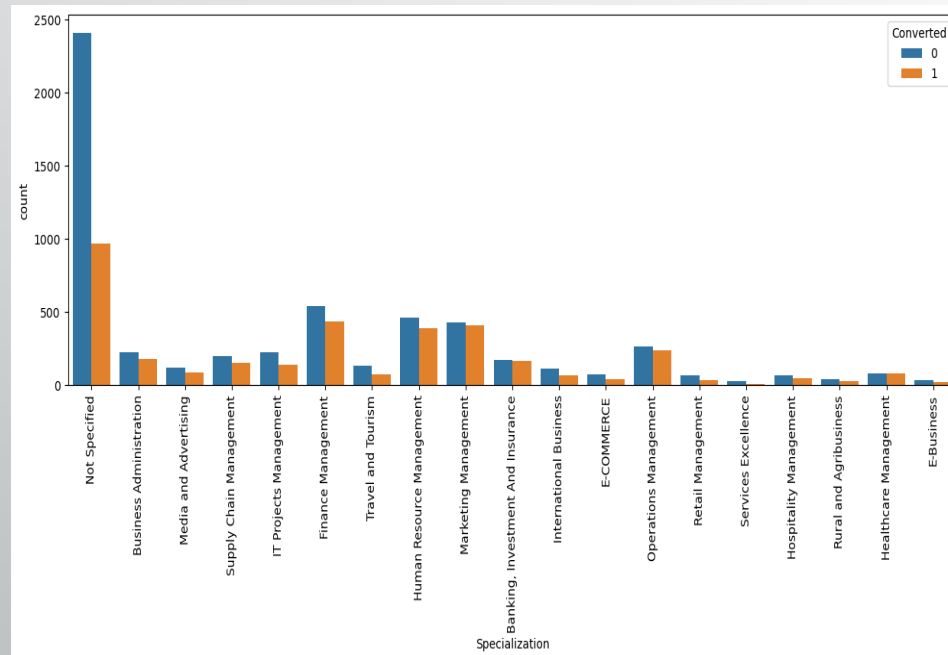
Most of the leads have no information about specialization marketing management ,human resources management has high conversion rate

Model evaluation (train)
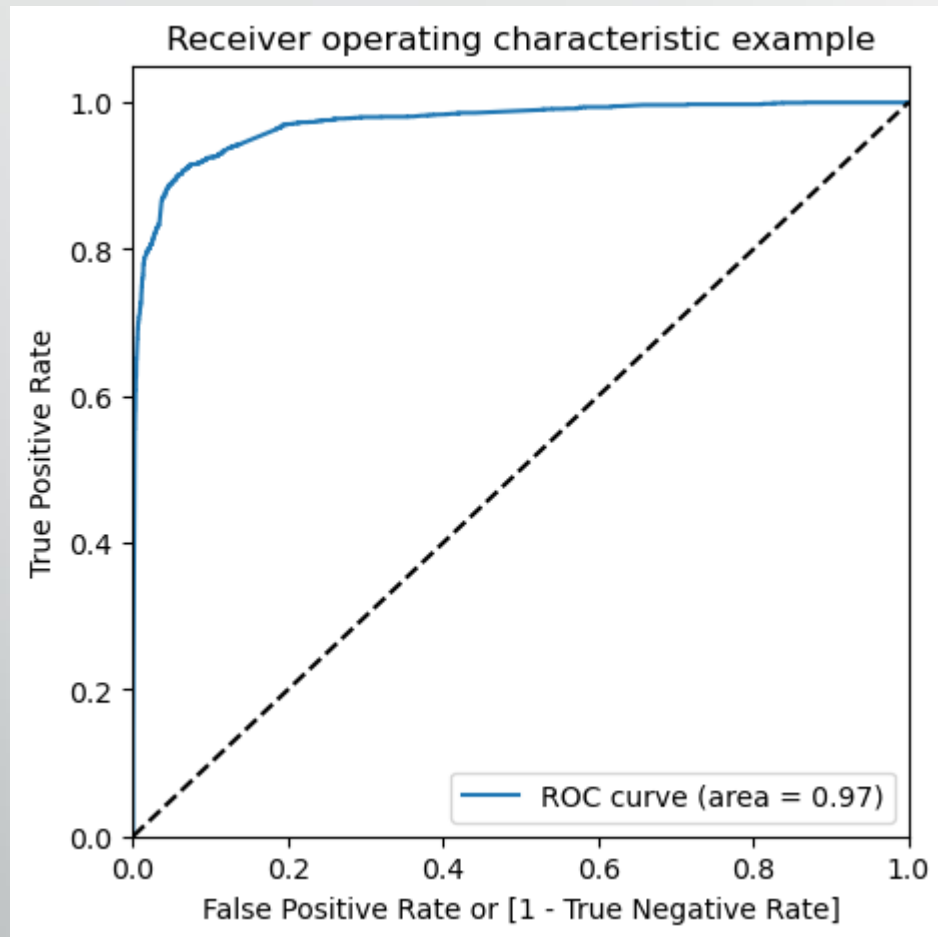
#The ROC Curve should be a value close to 1. We are getting a good value of 0.97 indicating a good predictive model.

From the curve above, 0.3 is the optimum point to

take it as a cutoff probability.
-accuracy=

The head of y_pred_final

	Prospect ID	Converted	Converted_prob	Lead_Score
0	5098	0	0.042068	4
1	4535	1	0.637697	64
2	3190	1	0.969043	97
3	7457	0	0.003775	0
4	6974	1	0.992921	99

	Prospect ID	Converted	Converted_prob	Lead_Score	final_Predicted
0	5098	0	0.042068	4	0
1	4535	1	0.637697	64	1
2	3190	1	0.969043	97	1
3	7457	0	0.003775	0	0
4	6974	1	0.992921	99	1