Lead Score Case Study

-Saurav Dash
-Saurabh Singh

Agenda

- Problem Statement
- Treating the data
- Key Specialization
- SMS! Not email
- Not just google
- Optimal cut-off
- Model Validation
- Inference

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

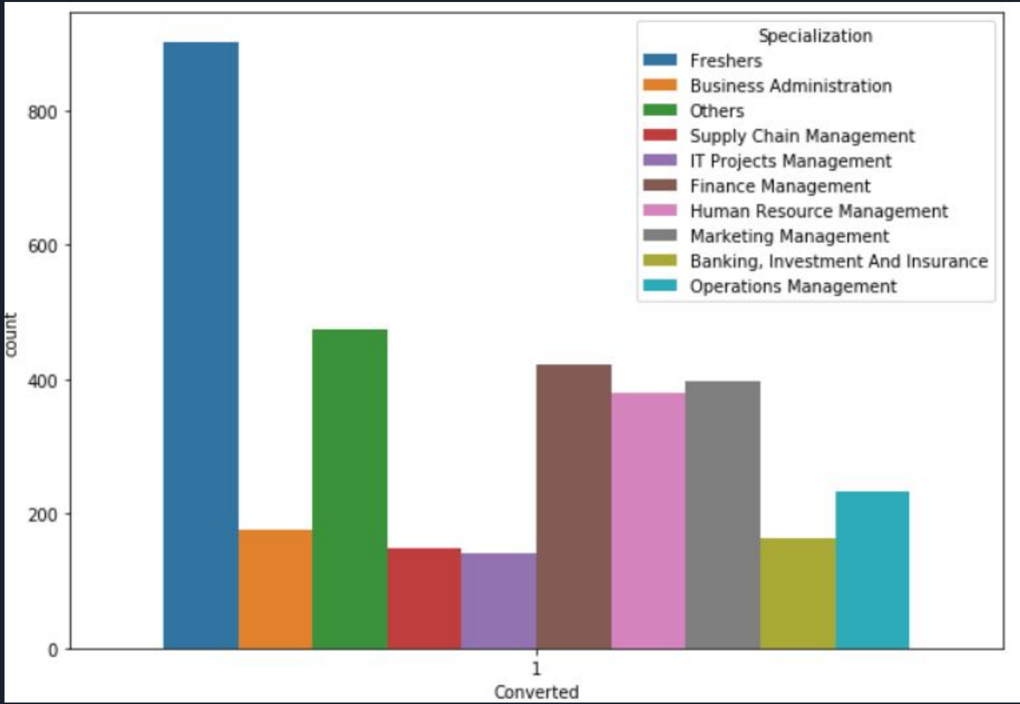
Treating the data

1. There were a lot of default values like 'Select' which might have occured due to any error from their side, their intent or the field was not applicable for that particular lead. So to stay on the safe side, we treated them as null values.

2. Many of the variables like specialization, lead source, last activity, last notable activity had many categories occuring in low frequencies.So we clubbed them into 1 category so as to avoid high number of dummies.

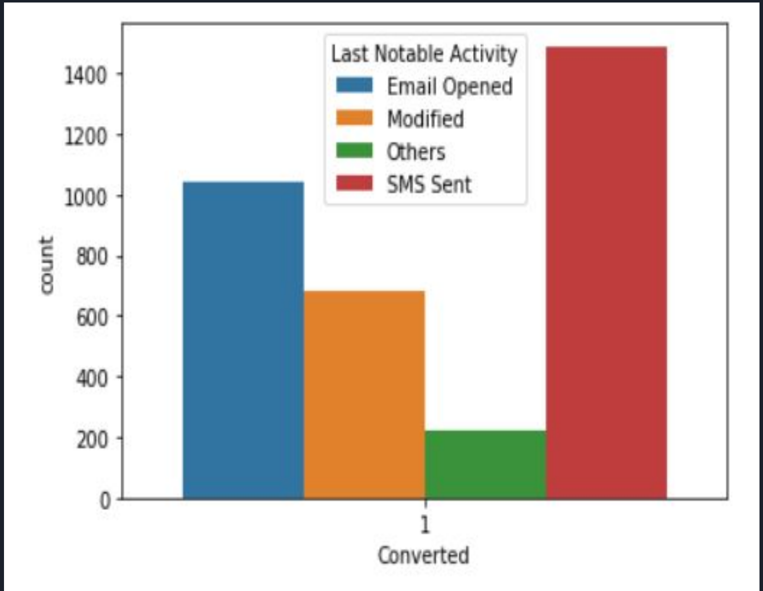
3. Outliers were present in the data. But their effect was not significant enough to distort our analysis, hence we skipped the step.

Key Specialization

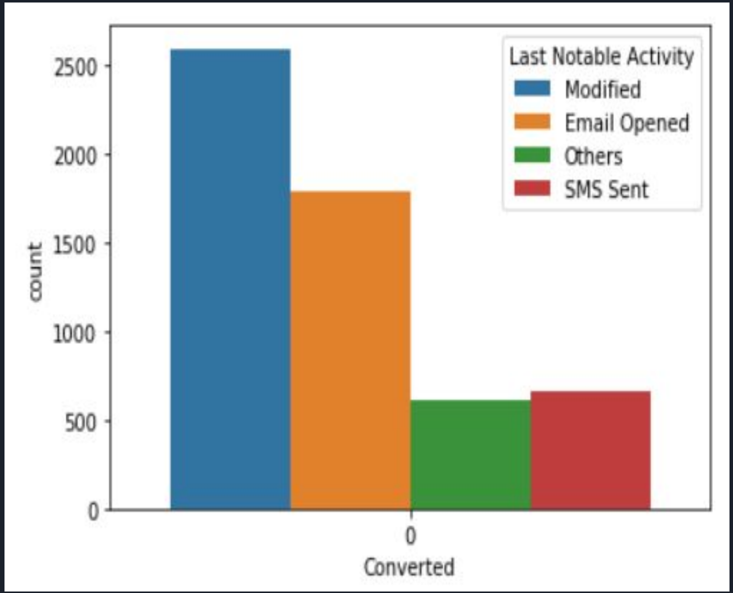


From the group of leads who were converted, we observe apart from freshers, leads with core specialization subject in masters tend to convert more.

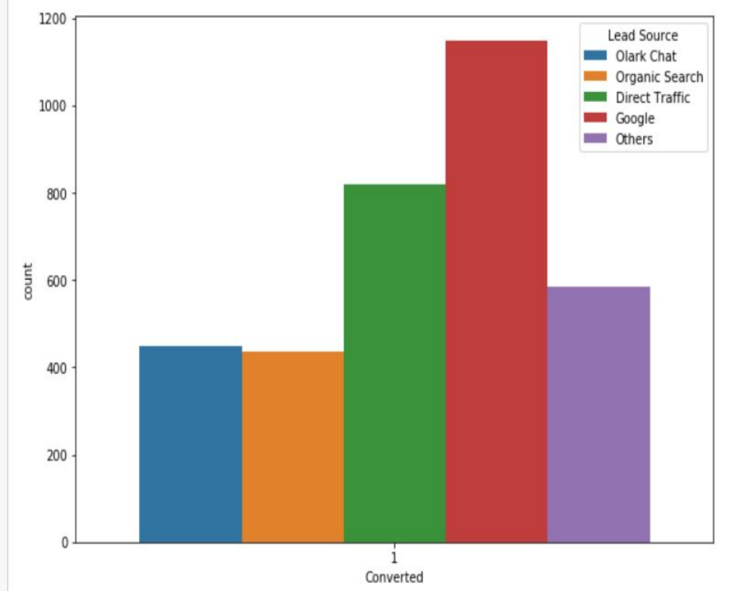
SMS! Not email.



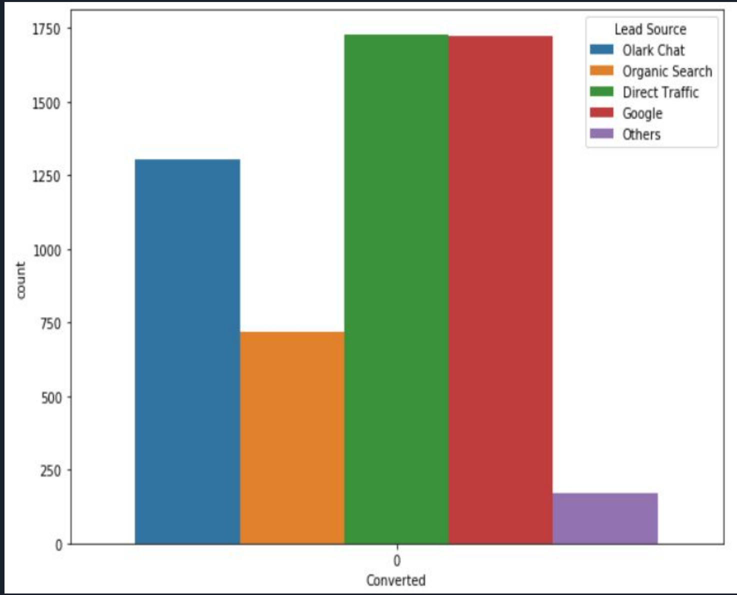
We observe that activity through SMS seemed to have higher chances of conversion. Hence equal importance is to provided to both email and sms category



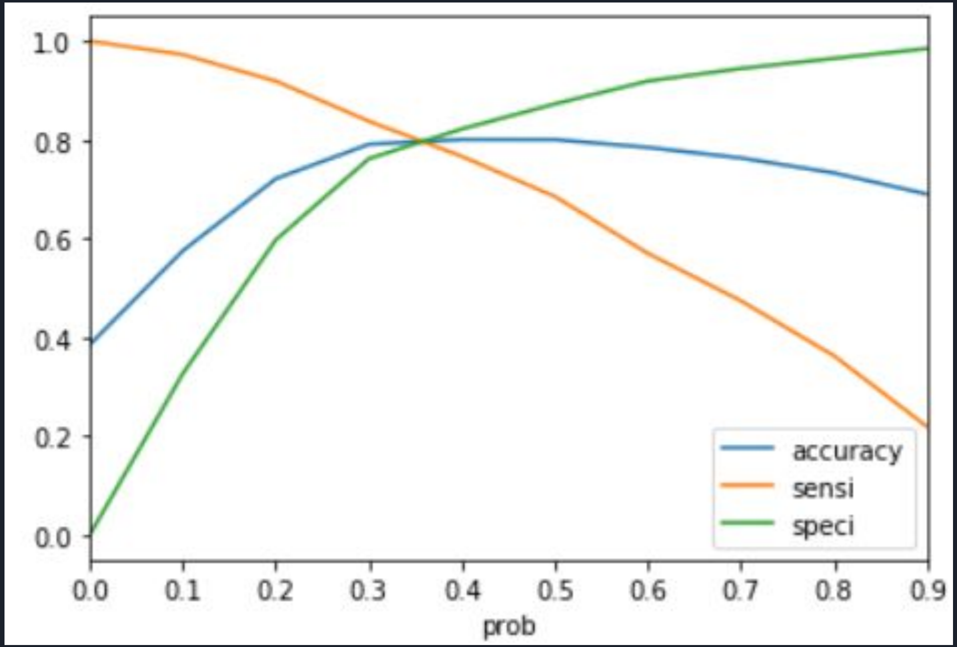
Not Just google.



Similarly importance needs to be given to leads coming from other sources like facebook,bing,live chats,click2call etc



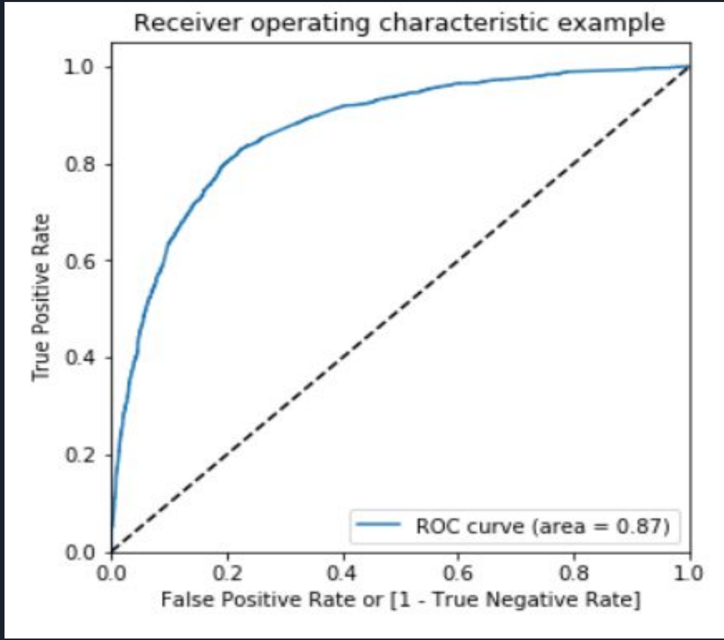
Optimal Cut-off



The point of intersection seems to be the optimal cut-off for the prediction keeping the sensitivity high (about 80%)

Here the value considered is 0.3, keeping some bias towards sensitivity

Model Validation

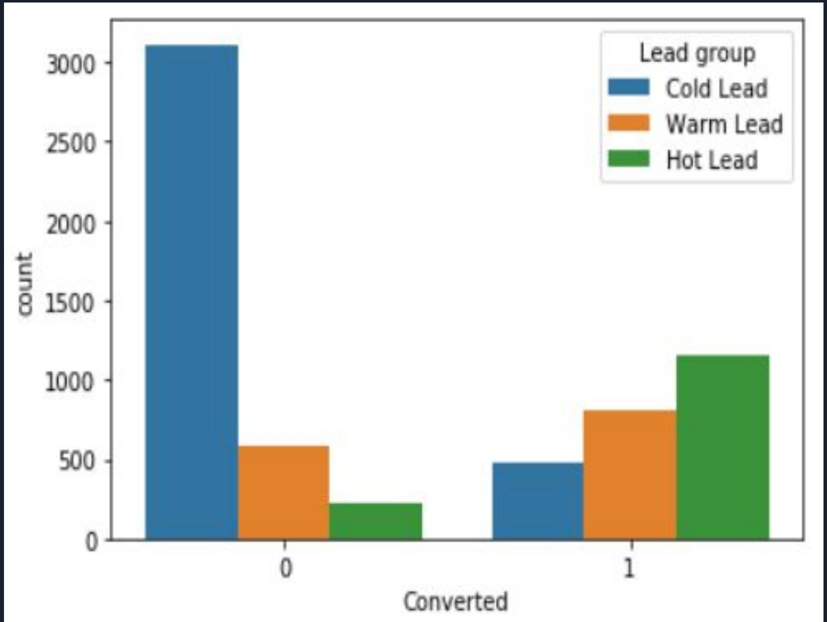


The tradeoff between sensitivity and specificity is quite good. Our problem was to increase the sensitivity which is being observed here.
The curve is towards the left corner border, hence confirming the accuracy of our test

Hence we obtained some good scores percentages

	Accuracy	Sensitivity	Specificity	Precision
Train Data	79.10%	83.76%	76.18%	68.78%
Test Data	78.22%	81.5%	76.29%	66.25%

Inference



Hence we categorise 3 kinds of leads from the lead scores obtained.

1. Hot leads (Scores above 70%) - An aggressive approach is needed for this category.A separate team can be assigned for this category.

2. Warm lead (Scores between 30% -70%) - Leads especially of core specialization subjects can be targeted first.

3. Cold lead (Scores less than 30%) - After meeting the deadlines, proper lead profiling can be done on this category to understand their interest and a better pitch call

Thank You