Lead Scoring Case Study

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:

Approach

- Uploading the data

- Read and understand the data

- Data Cleaning

- EDA

- Splitting the data in test & train dataset

- Prepare the data modelling

- Model Building

- Model Evaluation

- Summary

Data Sourcing and Cleaning

- Read the data from CSV

- Clean the data

- Remove unwanted Columns

- Imputing Null Values

- EDA

Model Building

- Feature selections using RFE

- Determined optional model using Logistic Regression

- Calculated accuracy, sensitivity, specificity, & evaluate model
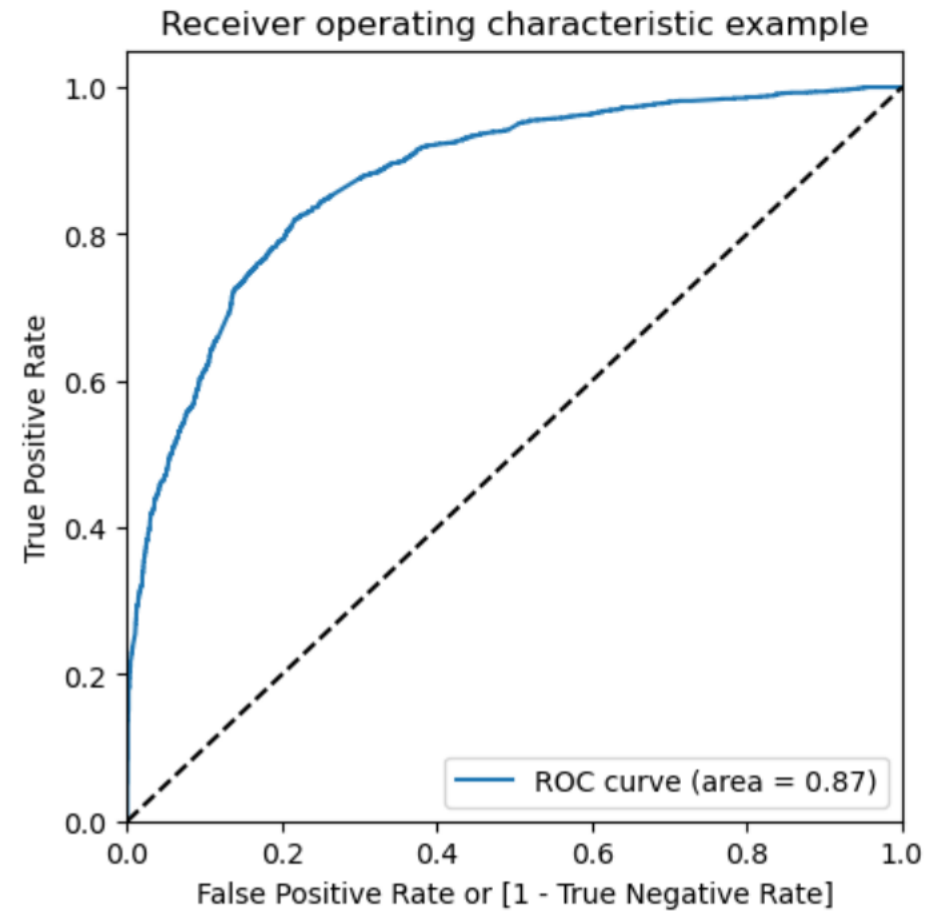
Variables impacting the conversion rate

- Total Visits

- Total time spent on website

- Lead source

- Lead origin

- Referrals

Model Evaluation sensitivity & specificity on train data set
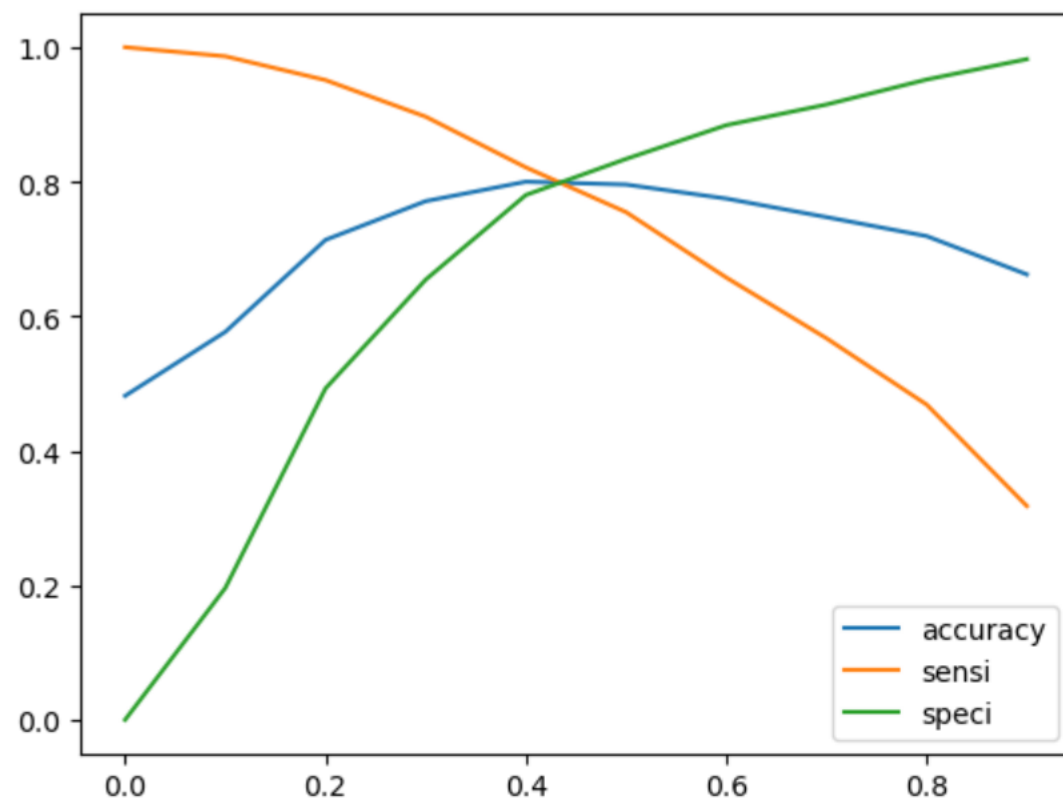
Receiver Operating characteristic Example

Result

- Accuracy, Sensitivity and Specificity values of training and test set are close to training set

- Accuracy, Sensitivity and Specificity values of training set are 79%,82%,76% Respectively

- Accuracy, sensitivity & Specificity values of test are 78%,81%,76% Respectively

- Conversion rate for Train & Test Dataset is 82.7% & 80.8% Respectively

- We have done the prediction on the test set using cut off threshold from sensitivity & specificity metrics.

Summary

- There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.
- First, sort out the best prospects from the leads you have generated. 'TotalVisits' , 'Total Time Spent on Website' , 'Page Views Per Visit' which contribute most towards the probability of a lead getting converted.
- Then, You must keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies. Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information or courses that suits best according to the interest of the leads. A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects.
- Focus on converted leads. Hold question-answer sessions with leads to extract the right information you need about them. Make further inquiries and appointments with the leads to determine their intention and mentality to join online courses.